Spring 2015

Transport studies in graphene-based materials and structures

Jiuning Hu
Birck Nanotechnology Center, Purdue University

Graduate School Form 30
(Revised 08/14)

PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By Jiuning Hu

Entitled
Transport Studies in Graphene-based Materials and Structures

For the degree of Doctor of Philosophy

Is approved by the final examining committee:

YONG P. CHEN

MARK S. LUNDSTROM

MINGHAO QI

XIULIN RUAN

To the best of my knowledge and as understood by the student in the Thesis/Dissertation Agreement, Publication Delay, and Certification/Disclaimer (Graduate School Form 32), this thesis/dissertation adheres to the provisions of Purdue University's "Policy on Integrity in Research" and the use of copyrighted material.

YONG P. CHEN

Approved by Major Professor(s): _____

Approved by: Michael R Melloch 04/17/2015

Head of the Department Graduate Program Date

TRANSPORT STUDIES IN GRAPHENE-BASED MATERIALS AND STRUCTURES

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Jiuning Hu

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

May 2015

Purdue University

West Lafayette, Indiana

Dedicated to my family.

ACKNOWLEDGMENTS

I would like to thank Prof. Yong P. Chen for offering me the opportunity and providing the platform to focus on interesting research topics. I also want to thank Prof. Xiulin Ruan (and his group members Dr. Bo Qiu, Yan Wang and Ajit Vallabhaneni) for many valuable technical support and comments for the projects of simulations. We have a wonderful crew on low temperature transport experiments. Most of the technical things that I have learned after I switched to experimental projects are from the lab members, Dr. Jifa Tian, Dr. Tailung Wu, Wonjun Park, Luis Jauregui, Jack Chung and Dr. Helin Cao. I want to specially thank Wonjun Park for being working together during the night when I was doing graphene flake hunt and transfer. I spent a fair amount of time in machine shop to build the flake transfer stage and VTI probe with loading lock. Dr. Tailung Wu provided tremendous support on system design, material selection, assembly and test. It is a wonderful experience to work with our outstanding lab members. I would like to thank Prof. Chin-Teh Sun and his student Niat Rahman for sharing and assisting on the hot press equipment. I also want to thank Ju-Hyun Park, Glover Jones and Tim Murphy for the assistance on experiments at National High Magnetic Field Laboratory. I want to thank the NIST staff David Newell, Curt Richter, Joseph Stroscio, Nikolai Klimov, Son Lei, Randolph Elmquist, Yanfei Yang and Qing Li for great support during my stay in NIST.

My deepest expression of appreciation goes to my parents, sisters and wife, for their unconditional love and supports.

TABLE OF CONTENTS

Page

Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

Figure Page

Figure Page

ABSTRACT

Hu, Jiuning Ph.D., Purdue University, May 2015. Transport studies in graphene-based materials and structures. Major Professor: Yong P. Chen.

Graphene, a single atomic layer of graphite, has emerged as one of the most attractive materials in recent years for its many unique and excellent properties, inviting a broad area of fundamental studies and applications. In this thesis, we present some theoretical/experimental studies about the thermal, electronic and thermoelectric transport properties in graphene-based systems. We employ the molecular dynamic simulations to study the thermal transport in graphene nanoribbons (GNRs) exhibiting various properties, including chirality dependent thermal conductivity, thermal rectification in asymmetric GNRs, defects and isotopic engineering of the thermal conductivity and negative differential thermal conductance (NDTC) at large temperature biases. We prove a theorem on the existence of NDTC in general one-dimensional diffusive thermal transport. We synthesis graphene composites and characterize their electric and thermal properties. Their electrical conductivity follows the percolation theory. We use 3-w method to measure the thermal conductivity and find that their thermal conductivities can be tuned by the graphene filler concentration. We build a micro-manipulator to accurately align and transfer two-dimensional materials, e.g., graphene and boron nitride (BN). We then fabricate the stacked BN/grapnene/BN/graphene/BN systems with isolated metal contacts to each graphene layer, to study the counterflow thermoelectric transport and Coulomb drag. In the last we present our theoretical considerations about the transport properties

of multilayer systems with interlayer Coulomb interactions, using phenomenological arguments, Drude-like models and Boltzmann transport formalism, and discussed the potential for the interlayer interaction to enhance the thermoelectric figure of merit.

1. INTRODUCTION

Graphene, a monolayer of graphite and two dimensional honeycomb lattice of sp^2 bonded carbon [1], is the building block of most carbon based nanomaterial, such as carbon buckyballs, nanotubes and graphite, as shown in Fig. 1.1. Since its famous Scotch-tape-isolation in 2004 [2], graphene and related research is greatly boosted not only in condensed matter physics, but also in many other branches of physics (e.g., optics and biophysics) and science (e.g., chemistry and material science). Even a brief overview of the graphene related research can immediately convince us the vast versatility of graphene in many aspects: unconventional quantum Hall effect, mimicking quantum electrodynamics, building super strong graphene composites, sensitive atom/molecule detectors, extreme high speed electronics, etc.

There are a few reasons for its versatility. First, from the material point of view, the building element of graphene, carbon (the 15th most abundant element in the earth's crust), although has more than 100 times less earth crustal abundance than another glorious element of silicon, it indeed composes the largest constitution of the materials produced on earth (more than 1000 times larger than silicon). In addition, the generation of graphene crystal is relatively easy with looser requirements than most other crystals. Recent studies show that graphene can be produced from any materials containing carbon atoms [3]. Thus, the raw materials to produce graphene are virtually everywhere accessible. These days, large scale production of graphene from chemical vapour deposition [4] and epitaxial growth [5] are already applied to various graphene electronics [6]. Second, graphene has exhibited fruitful physics.

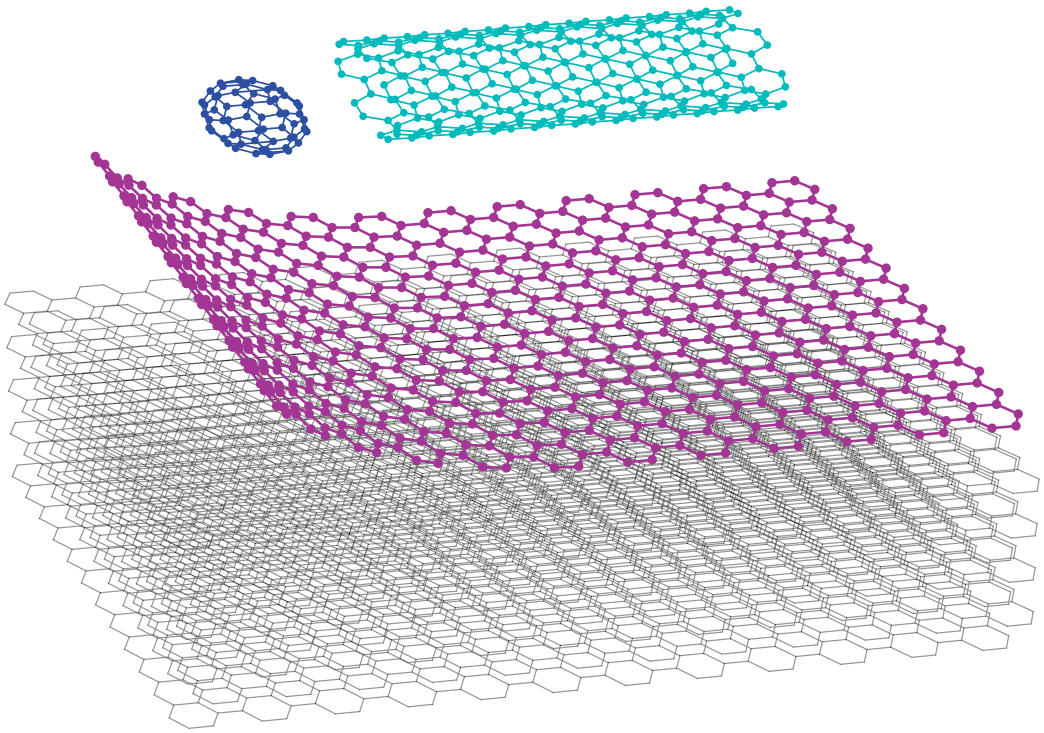

Fig. 1.1. Graphene (magenta, to be exfoliated from graphite) as the building blocks of buckyballs (blue), nanotubes (cyan) and graphite (gray).

Generally, in the field of condensed matter physics, much attention has been given to the exceptional and unique electronic properties of graphene [2, 7–11] uncovered in the past decade. Most excitingly, graphene could be a test bed for fancy physics, e.g., table top high energy physics [12] and exciton condensation, which requires very clean graphene system. Third, graphene can actively involve many physical and chemical processes to create composite materials. Due to its true two-dimensional (2D) nature, graphene exhibits extremely low percolation value [13]: graphene is able to drastically alter the matrix electrical conduction properties with tiny filling. Standalone graphene is one of the strongest materials in the world [14]. By adding graphene,

the mechanical strength of graphene composites can be enhanced, as demonstrated in graphene paper [15]. Fourth, the 2D nature of graphene enables us to easily pattern graphene planar nano structures. These nano structures have been considered as important elements in future carbon-based nanoelectronics. For example, it has been shown that many of the electronic properties of these patterns may be tuned by its width or edge structures [16–19]. This nano patterning (along with other materials) is not only happening in-plane, but also extending in the out-of-plane direction, e.g., graphene and boron nitride (BN) hybrid systems [20]. Fifth, graphene can be in stacking with other materials (e.g., BN and superconducting films) to study interaction physics, e.g., Coulomb drag and exciton condensation.

The main theme of the thesis is studying the energy and carrier transport properties in graphene-based structures and materials. Both theoretical and experimental tools are used to carry out those studies. Among them, classical molecular dynamics (MD) simulation is a convenient tool to study the lattice dynamics of graphene nanoribbons (GNRs) that consist of up to few thousands carbon atoms in our simulations. The size of those GNRs is around 10 nm. It is very challenging to perform measurements on such small systems. However it is much easier to perform simulation. We study the (lattice) thermal transport in GNRs at various temperatures, chirality, shape and compositions. A few results [21–28] are obtained, and we expect that they open possibilities of using graphene nanostructures for nanoscale thermal management. On experimental side, two limiting situations are examined. One is the "dirty" limit of graphene composite in which the thermal conductivity are measured using 3ω [29] method. Another is the "clean" limit that high mobility graphene samples are pursued. Graphene on local BN substrate is demonstrated to have high mobility [30, 31] compared to that on SiO_2/silicon substrate, with much less effort than fabricating suspended graphene. More importantly, we can build many kinds

of graphene-based heterostructures, e.g., two graphene layers separated by a thin BN to perform Coulomb drag measurement and study excitonic physics. A condensed introduction to some important aspects is given in the following.

1.1 Molecular dynamics

Classical MD simulation conducted in this thesis is one of the most popular atomistic simulations and it is focusing on the atomic movements using Newtonian dynamics for interacting many-body systems. The many-body potentials are generated using empirical formulas that can capture the major features of the potential between atoms. For example, the reactive empirical bond-order (REBO) potential used in this thesis is based on the Tersoff's formalism and lately extended to adaptive intermolecular REBO potential [32, 33], and can describe the covalent bondings for hydrocarbon systems. By imposing appropriate initial and boundary conditions, the dynamics of atoms can be either deterministic or stochastic. In our simulations, we use the Nosé-Hoover thermostats [34, 35] thermostats which is deterministic to assign local temperatures. The heat current can be calculated to deduce the thermal conductivity/conductance.

Although it is a convenient tool, it suffers from limitations. First, the interatomic potential is of quantum mechanical nature and is often approximated using empirical equations. Secondly, the quantum mechanical statistical nature of phonons cannot be encoded in the simulation process. Instead, a post hoc quantum correction is carried out for the local temperatures.

1.2 Graphene heterostructures

The 2D nature of graphene allows it not only to be fabricated into various sizes and shapes, but also to be stacked with other materials in the direction normal to the graphene plane. The stacking of graphene with other 2D materials (e.g., BN and niobium selenide) can be achieved by using the flake transfer technique [30, 36–38].

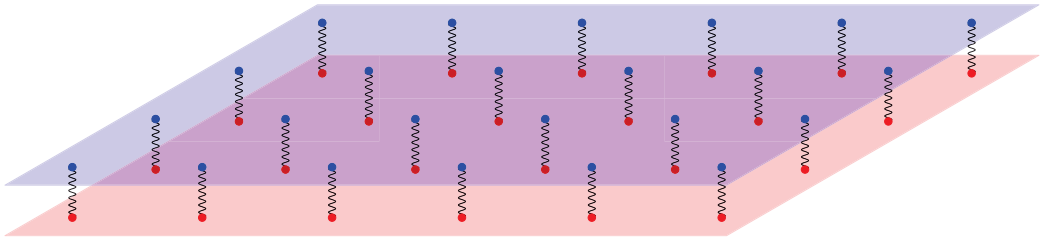

Fig. 1.2. Interlayer excitons. The wiggle lines represent the binding between interlayer carriers.

For two graphene layers separated by a thin insulator and one layer is p-type and the other is n-type, the intriguing exciton condensation [39] is predicted to occur, though the predicted transition temperature spans a wide range [40–49]. The interlayer electron-hole pairs in Fig. 1.2 are represented by a pair of red and blue dots residing in the bottom read and top blue layers with bonding of wiggle lines. Unlike excitons excited by light in a single material, these interlayer excitons in principle have infinite life time. The interlayer exciton condensation is only observed in coupled quantum wells [50] in magnetic field. Great experimental efforts have been devoted, but it remains an open question whether exciton condensation can be observed in graphene double layers. It is also unclear how the strong interlayer Coulomb interaction can affect the thermoelectric transport of excitons (if existed) and carriers. These issues will be partially addressed in the thesis.

1.3 Outline of this thesis

This thesis is organized as follows.

In chapter 2, we present our theoretical work on thermal transport in GNRs and its related systems. We use MD to simulate the thermal transport in various graphene nanostructures, such as rectangular, triangular and trapezoidal GNRs, GNRs with edge hydrogenation, and GNRs composed of carbon isotopes. We also extend the simulation beyond the linear response regime and find that there exists negative differential thermal conductance (NDTC) in short GNRs. We then analytically consider the existence of NDTC in general 1D diffusive thermal transport. To our surprise, there cannot exist NDTC in such systems when the temperature at one end is fixed, even though the thermal conductivity can be locally engineered. However, by introducing a temperautre dependent thermal contact resistance (TCR), there could exist NDTC. Our predictions could be useful to build the first device that experimentally exhibits NDTC effect.

In chapter 3, we present our experimental work on thermal and electrical transport in graphene composite systems. We synthesize graphene composite using chemically reduced graphene and polystyrene. The dried composite is then hot pressed into bulk plate for measurement. We prepare samples with different graphene filling concentrations. We develop a novel transfer method to place micrometer sized heater/sensor on composite surface to avoid damaging the composite materials during traditional lithography processes. The measured thermal conductivity is increasing with the filling fraction of graphene. The enhancement can be as high as 50% for 5 vol.% filling.

In chapter 4, Coulomb drag and counterflow thermoelectric transport measurements are performed in layered structures of BN/BLG/BN/BLG/BN (BLG stands

for bilayer graphene). The magnitude of the counterflow Seebeck coefficient exhibits a peak in the regime where two graphene layers have opposite sign of charge carriers. The maximum power factor is about $700\mu\mathrm{W/K^2cm}$ at room temperature. The counterflow Seebeck coefficient and power factor decrease approximately linearly with temperature from 300 K to 50 K. The measured interlayer Coulomb drag resistivity is negligibly small ($< 3\ \Omega$), compared to the intralayer resistivity, suggesting negligible impact of Coulomb drag on the counterflow thermoelectric transport.

In Chapter 5, we present our theories of carrier transport in multilayer systems in presence of strong Coulomb drag. We offer a phenomenological theory of how the presence of strong Coulomb drag is affecting the (thermoelectric) transport properties of double layer systems. For Coulomb drag resistivities that satisfying some specific relations, the (counterflow) transport in double (multiple) layer system could be dissipationless. The Drude-like model provides some insights about the conditions on the relaxation time of Coulomb scattering between carriers in different layers. The Boltzmann formalism is employed to study the Coulomb drag conductivity and its reciprocal relations. It is pointed out that the existence of interlayer excitons might break the reciprocal relations. Our theory is also generalized to a generic multi-layer system.

The thesis is summarized and comments on future directions are provided in chapter 6.

2. THEORETICAL STUDIES OF THERMAL TRANSPORT IN GRAPHENE NANORIBBONS

Graphene nanoribbons (GNRs), narrow strips of graphene with few or few tens of nm in width, are particularly interesting and have been considered as important elements in future carbon-based nanoelectronics. In addition to its electronic properties, the thermal properties of graphene are also of fundamental and practical importance. It has been demonstrated that graphene has a superior thermal conductivity [51]. This opens numerous possibilities of using graphene nanostructures for nanoscale thermal management.

In our simulations, we have used classical molecular dynamics (MD) based on the Brenner potential [32, 33]. The Brenner potential is a many-body potential for hydrocarbon systems that considers the coordination number of each atom. The anharmonic effects, causing the phonon scattering and consequently leading to finite value of thermal conductivity, have been automatically embedded in the Brenner potential. Our MD simulation objects are GNRs of typically 1-4 nm wide and 6-10 nm long.

We have observed that the thermal conductivity of GNRs depends on the edge-chirality and can be affected by defects [21]. We have also observed thermal rectification (TR) in asymmetric GNRs, where the thermal conductivity in one direction is significantly different from that in the opposite direction [21, 26]. TR has been experimentally observed in asymmetrically mass loaded nanotubes [52] and theoretically predicted in several other carbon nanostructures such as carbon nanohorns [53]

and carbon nanotube intramolecular junctions [54]. TR have potential applications in nanoscale thermal management such as on-chip cooling and energy conversion by controlling the heat transport, and is also fundamental in several recently proposed novel schemes of thermal circuits or information processing using phonons [55–58]. We also have studied the thermal transport in edge-hydrogenated and isotopically engineered graphene systems [24]. We have found that the thermal conductivities can be dramatically tuned by the isotope doping patterns.

Beyond the linear response of thermal transport, we have found that there exists negative differential thermal conductance (NDTC) in GNRs [25]. We also investigated the existence of NDTC in general 1D diffusive thermal transport.

2.1 Molecular dynamics simulation procedure

Fig. 2.1 shows the typical GNRs in our simulations. The carbon atoms denoted by triangles in Fig. 2.1 at the two ends of the GNR are placed in the Nosé-Hoover thermostats [34, 35] (obeying Eq. (2.1)) at temperatures $T_{\rm L}$ (left-pointing triangles) and $T_{\rm R}$ (right-pointing triangles) respectively. The absolute value of the temperature difference is $\Delta T = |T_{\rm L} - T_{\rm R}|$. The equations of motion for atoms in either the left or right Nos-Hoover thermostat are:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{p}_i = \mathbf{F}_i - \gamma\mathbf{p}_i, \quad \frac{\mathrm{d}}{\mathrm{d}t}\gamma = \frac{1}{\tau^2}\left[\frac{T(t)}{T_0} - 1\right], \quad T(t) = \frac{2}{3Nk_{\rm B}}\sum_i \frac{\mathbf{p}_i^2}{2m_i}, \tag{2.1}$$

where the subscript i runs over all the atoms in the thermostat, $\mathbf{p}_i$ is the momentum of the i-th atom, $\mathbf{F}_i$ is the total force acting on the i-th atom, γ and τ are the dynamic parameter and relaxation time of the thermostat, $T(t)$ is the instant temperature of the thermostat at time t, T_0 (=$T_{\rm L}$ or $T_{\rm R}$) is the set temperature of the thermostat, N is the number of atoms in the thermostat, $k_{\rm B}$ is the Boltzmann constant and m_i is the

mass of the i-th carbon atom. We used the 3$^{\rm rd}$-order prediction-correction method to integrate these equations of motion. The time step is 0.5 fs and the simulation is run for 10^7 time steps (the total MD time is 5 ns). The relaxation time τ is 1 ps. Typically, $T(t)$ can stabilize around the set value T_0 after about 0.5 ns. We did the time averaging of the temperature and heat current from 2.5 ns to 5 ns. The heat current injected to the left (right) thermostat is given by [54]

$$J_{\rm L(R)} = \sum_{i\in {\rm L(R)}} (-\gamma \mathbf{p}_i) \cdot \frac{\mathbf{p}_i}{m} = -3N\gamma k_{\rm B} T_{\rm L(R)}(t) \tag{2.2}$$

where $T_{\rm L(R)}(t)$ is the instant temperature for the left (right) thermostat.

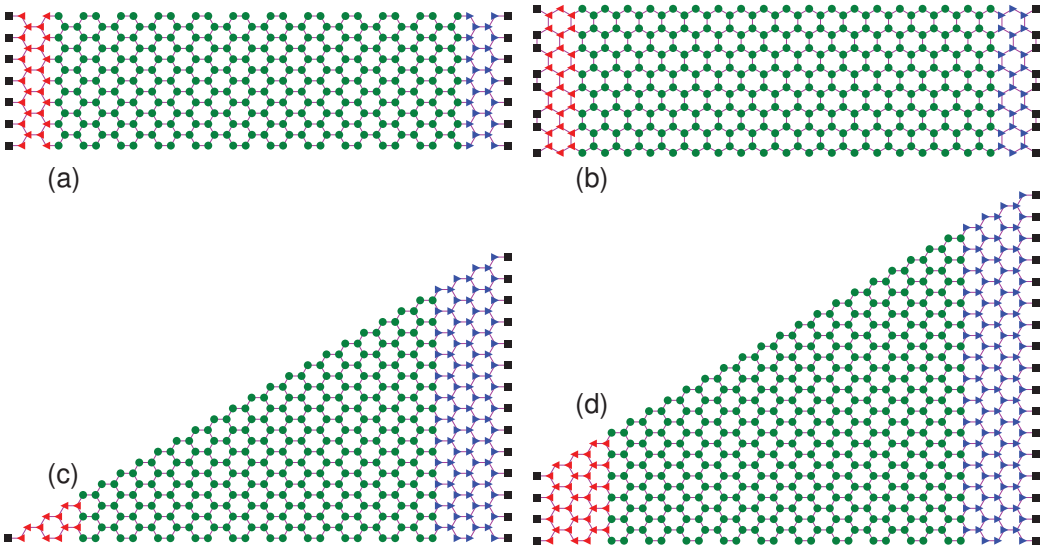

Fig. 2.1. Typical GNR structures: (a) rectangular armchair GNR, (b) rectangular zigzag GNR, (c) triangular GNR with armchair bottom edge and (d) trapezoidal GNR with armchair bottom edge. ■ denotes fixed boundary atoms. ◀ (▶) denotes atoms in the left (right) thermostat. ● denotes the remain atoms in the bulk.

In Fig. 2.1, the atoms denoted with squares at the ends are fixed to avoid the spurious global rotation of the GNRs in the simulation [59]. The atoms denoted with circles obey the Newtons law of motion:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{p}_j = \mathbf{F}_j, \tag{2.3}$$

where j runs over all atoms neither fixed nor in any thermostats.

Thermal conductivity is ideally defined as the limit $\kappa = \lim_{\Delta T \to 0} Jd/(\Delta T w h)$ according to Fouriers law, where d, w and h (=0.335 nm) are the length, width and thickness of the GNR respectively. In non-equilibrium MD simulation, however, this mathematical limit cannot be achieved due to the fluctuations in both the temperature difference ΔT and the resulting heat current J. If ΔT is too small to dominate its fluctuations, the calculation of thermal conductivity could become problematic. To address this problem, we defined $\alpha = \Delta T/(2T)$ where $T = (T_\mathrm{L} + T_\mathrm{R})/2$ and examined the thermal current as a function of α for the GNR in Fig. 2.1a. The result is shown in Fig. 2.2. We found that it is still in linear response region up to $\alpha = 10\%$, so we set $\alpha = 10\%$ in all our simulations that calculate the thermal conductivity or conductance. At steady state, in principle $J_\mathrm{L} + J_\mathrm{R} = 0$ because there is no energy accumulation. However, $J_\mathrm{L} + J_\mathrm{R}$ slightly fluctuates around zero. Therefore, the heat current and its error are defined as $J = (J_\mathrm{L} - J_\mathrm{R})/2$ and $\Delta J = |J_\mathrm{L} + J_\mathrm{R}|/2$ (corresponding to the error in the thermal conductivity $\Delta\kappa = \Delta J d/(\Delta T w h)$).

We corrected the temperature calculated from Eq. (2.1) by considering the phonon occupation number and density of states (DOS) of the 2-dimensional graphene lattice. We only consider phonons of three acoustic branches and assume linear dispersion with phonon sound velocities [60] $v_\mathrm{LA} = 19.5$ km/s, $v_\mathrm{TA} = 12.2$ km/s and $v_\mathrm{ZA} = 1.59$ km/s. Therefore, the phonon DOS is $D(\omega) = 3\omega/(2\pi n v^2)$, where v is the effective phonon sound velocity which is calculated from $3v^{-2} = v_\mathrm{LA}^{-2} + v_\mathrm{TA}^{-2} + v_\mathrm{ZA}^{-2}$ and n is the

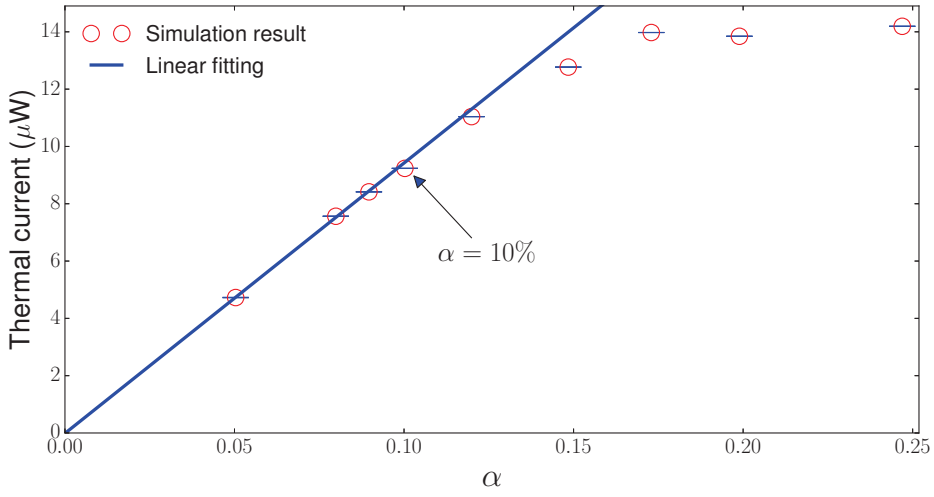

Fig. 2.2. Thermal current versus (solid line) and its linear fit (dashed line).

number of carbon atoms in unit area of graphene. The phonon energy per carbon atom at temperature T from the Debye model should equal to the kinetic energy per carbon atom $3k_BT_{\rm MD}$, i.e.,

$$\int_0^{\omega_{\rm D}} D(\omega)n(\omega,T){\rm d}\omega = 3k_{\rm B}T_{\rm MD}, \quad n(\omega,T) = \frac{1}{e^{\hbar\omega/k_{\rm B}T}-1}, \tag{2.4}$$

where $\omega_{\rm D} = k_{\rm D}v$ ($k_{\rm D} = \sqrt{4\pi n}$) is the Debye frequency, $n(\omega,T)$ is the phonon occupation number and $T_{\rm MD}$ is the classical MD temperature. The scheme of quantum correction of temperature can be obtained from Eq. (2.4):

$$T_{\rm MD} = \frac{2T^3}{T_{\rm D}^2}\int_0^{T_{\rm D}/T}\frac{x^2}{e^x-1}{\rm d}x \tag{2.5}$$

where $T_{\rm D} = \hbar\omega_{\rm D}/k_{\rm B}$ is the Debye temperature (322 K). It is worthy to note that the zero-point-energy is not included in the phonon occupation number because the corresponding temperature is one third of the Debye temperature (107 K), which

means that there is no corresponding quantum corrected temperature if the classical MD temperature is below 107 K. Fig. 2.3 shows the difference after introducing the zero-point-energy during the quantum correction. In all of the following MD results, the temperature axis represents quantum corrected temperature.

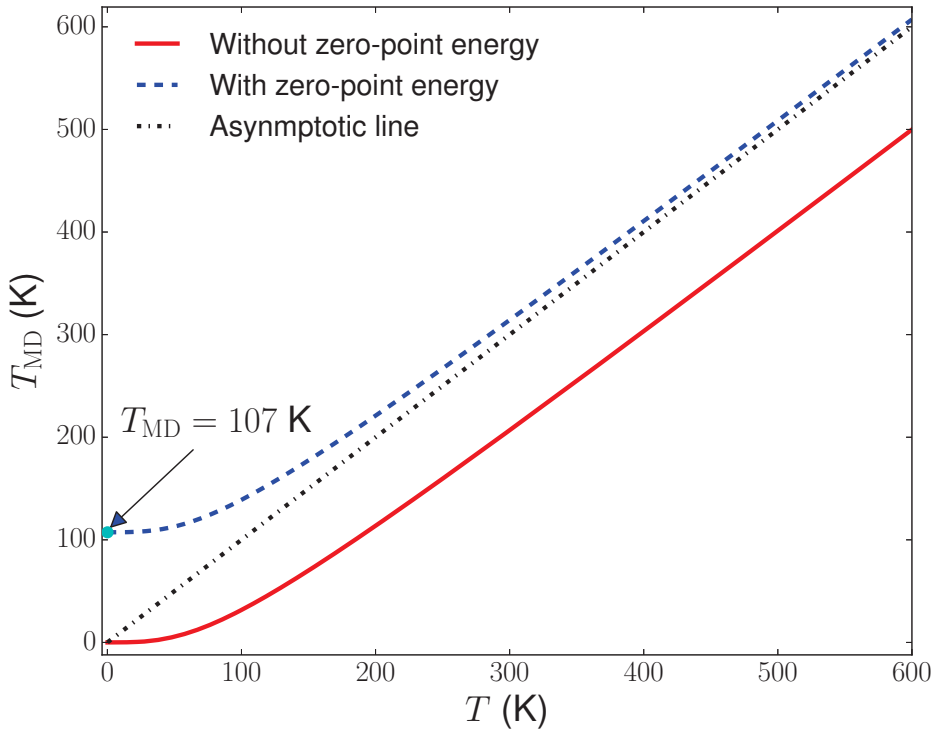

Fig. 2.3. Quantum correction of the MD temperature. The solid (dashed) line is the result without (with) the zero-point-energy included. The dotted is the asymptotic line at high temperature [22].

2.2 Thermal conductivity and thermal rectification of graphene nanoribbons

We have calculated the thermal conductivity of the rectangular GNR in Fig. 2.1a. The solid line in Fig. 2.4 shows the thermal conductivity of this 5.7 nm long and 1.5 nm wide GNR as a function of temperature. The thermal conductivity is around 1700 W/m-K at 400 K (Fig. 2.4), on the same order of magnitude with the experimental measured value ($\sim$ 600-5000 W/m-K) of graphene [61–65].

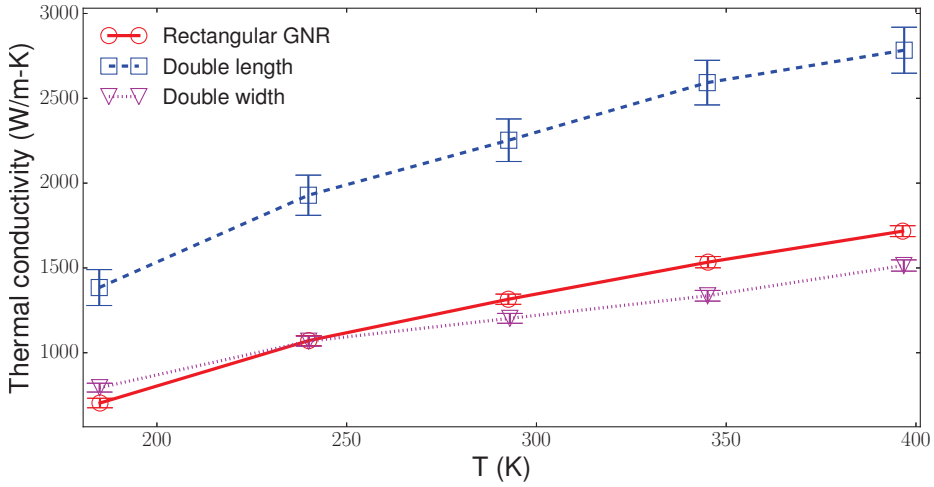

Fig. 2.4. Thermal conductivity of the GNR in Fig. 2.1a (solid line) and that of the GNR with double length (dashed line) and width (dotted line) only.

The effect of edge chirality on the thermal conductivity in rectangular GNRs is also investigated. The chirality of GNRs (see the right inset of Fig. 2.6 for the definition of chirality degree) is defined according to the edge parallel to the long direction of the GNR (different from the convention for CNTs). In Fig. 2.5, the thermal conductivity is plotted as a function of temperature for both zigzag (dashed

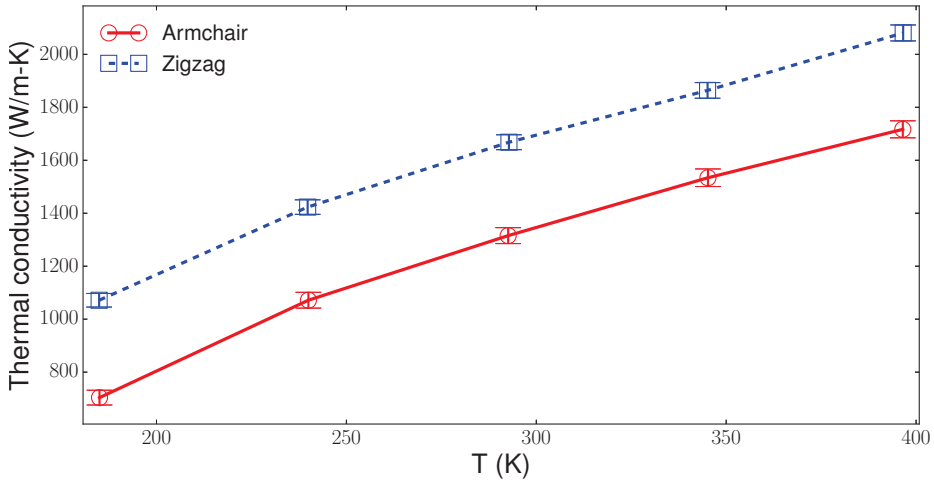

Fig. 2.5. Thermal conductivity of armchair (Fig. 2.1a) and zigzag (Fig. 2.1b) GNRs.

line) and armchair (solid line) edged GNRs. We show that the thermal conductivity of zigzag GNR (dashed line) is $20 - 50\%$ larger than that of the armchair GNR (solid line). Recent work by Jiang et al. obtained qualitatively similar results as ours using an approach based on ballistic phonon transport [66]. Similar effects have also been studied for CNTs, but there have been no generally accepted agreement on the preferred chirality of CNT for heat conduction [67, 68]. CNTs are periodic in the azimuthal direction and they can be considered as GNRs with infinite width. All phonon modes can propagate in CNTs. However, in GNRs with finite width the phonon modes propagating along the heat flow direction dominate. We speculate that the difference between the thermal conductivity of armchair and zigzag GNRs is mainly due to the different phonon scattering rates at the armchair and zigzag edges and the effect of the finite size of GNRs. Fig. 2.6 shows the thermal conductivity as a function of the chirality angle degree at the temperature of 400 K. Two peaks of the

thermal conductivity at chirality angles of 0° (corresponding to armchair edge) and 30° (zigzag edge) can be clearly identified. The peak at 30° is higher than that at 0°. We only need to study the chirality angle from 0° to 60° due to the 6-fold rotational symmetry of graphene. The GNRs with chirality angle not equal to integer multiples of 30° have irregular edges. Phonons can be strongly scattered by these irregular edges, likely resulting in the relatively low thermal conductivity at these angles (thus the two peaks at 0° and 30°).

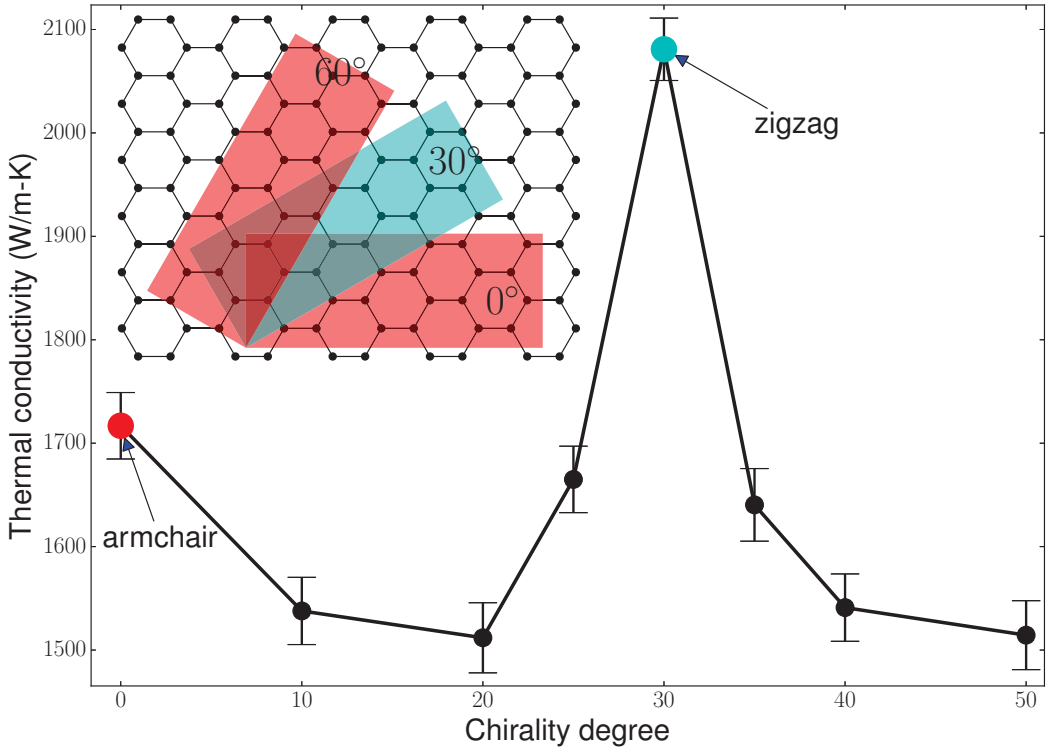

Fig. 2.6. Thermal conductivity as a function of chirality angle degree defined in the inset.

We studied the effect of the finite size of GNRs on their thermal conductivities. In Fig. 2.4, When the length of this GNR is doubled (the width remaining unchanged), the calculated thermal conductivity almost doubles (dashed line). It is probable that our calculated thermal conductivity is limited by the finite length of GNRs, which is consistent with the large phonon mean free path (MFP) in graphene extracted from the experiment (775 nm) [69]. The size of GNRs in this study is only up to 30 nm. Therefore, our calculated thermal conductivity is not corresponding to the value for graphene of macroscopic size. In addition, we have also found that the calculated thermal conductivity remains nearly the same after doubling the width of GNRs (the length remains unchanged), as indicated by the dash-dotted line of Fig. 2.4. As we see from Fig. 2.4, the thermal conductivity monotonically increases with temperature (T), in the range we studied (180-400 K), for GNRs up to 10 nm long. Our results are consistent with a recent theory by Nika et al. on small graphene flakes (e.g., 5 m long) where the important roles of phonon scattering by graphene edges in the temperature dependence of thermal conductivity have been discussed [70].

We also investigate the length dependence of the thermal conductivity of rectangular GNRs with armchair or zigzag long edges. We find that the GNRs with zigzag long edges have larger thermal conductivity than GNRs with armchair long edges. Fig. 2.7 shows the calculated thermal conductivity of GNRs with armchair (solid line) and zigzag (dashed line) long edges as a function of their length at $T = 400$ K. The thermal conductivity of GNRs increase with their length, which is consistent with the result of Fig. 2.4 and phonon MFP limited value of thermal conductivity.

We have studied thermal conductivity of triangular GNRs of Fig. 2.1c and trapezoidal GNRs of Fig. 2.1d and found significant thermal rectification, as shown in Fig. 2.8. In calculating the thermal conductivity of asymmetric GNRs, the width (w) is taken as the width at the middle of GNRs. For these structurally asymmetric

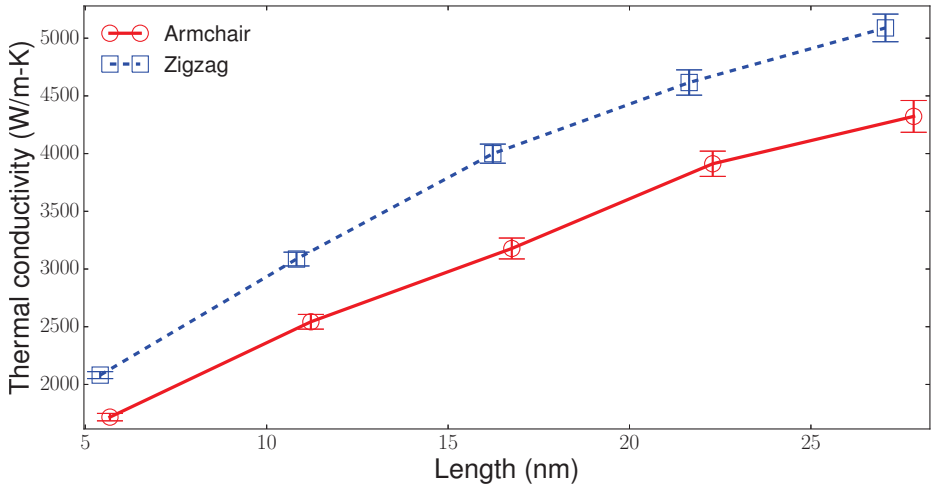

Fig. 2.7. Thermal conductivity of rectangular GNRs with armchair (solid line) and zigzag (dashed line) long edges versus their length at 300K [22].

GNRs, the thermal conductivity from the narrower (N) to the wider (W) end ($\kappa_{\mathrm{N}\to\mathrm{W}}$) is larger than that from the wider to the narrower end ($\kappa_{\mathrm{N}\to\mathrm{W}}$). The thermal rectification factor, defined as $\eta \equiv (\kappa_{\mathrm{W}\to\mathrm{N}} - \kappa_{\mathrm{N}\to\mathrm{W}})/\kappa_{\mathrm{N}\to\mathrm{W}}$, is as large as 120% (at $T = 180$ K), as shown in the inset of Fig. 2.8. The trapezoidal GNRs have higher thermal conductivity than triangular GNRs, possibly due to a combination of the geometry and size dependence of thermal conductivity. For the symmetric rectangular GNR, the thermal conductivity from the left (L) to the right (R) end is almost the same as that from the right to the left end and thus it exhibits nearly zero thermal rectification factor. Actually the nearly zero value of η is most probably from the numerical uncertainty and fluctuations related to MD simulation. The temperature dependence of the thermal conductivity of the rectangular or trapezoidal GNRs is found to be similar to that of the rectangular GNRs.

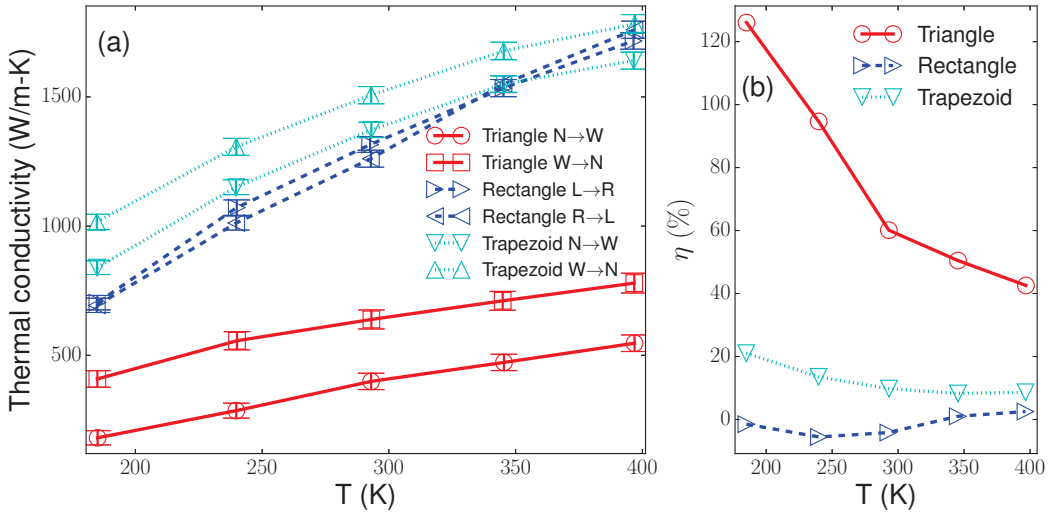

Fig. 2.8. Thermal conductivity of a rectangular GNR of Fig. 2.1a, a triangular GNR of Fig. 2.1c and a trapezoidal GNR of Fig. 2.1d. The inset shows the corresponding thermal rectification factor η as a function of temperature.

Previous studies of thermal rectifications (TR) [52–55] have suggested that TR originates from the interplay between structural gradient and lattice nonlinearity. A similar mechanism can underlie the TR in our asymmetric GNRs. We have investigated several geometric variations of the triangular GNRs to search for the structure with largest thermal rectification. To make better comparison, the solid lines in Fig. 2.9a and Fig. 2.9b are the calculated thermal conductivity of asymmetric triangular GNR shown in Fig. 2.1c. This particular structure has an armchair bottom edge and a 30° vertex angle (defined in Fig. 2.1a). Fig. 2.9b shows the thermal rectification of the GNRs with different vertex angles but the same bottom edge length and chirality (armchair). The dashed lines (dotted lines) are for the GNRs with vertex angle of 45° (60°). Compared with the GNR with vertex angle of 30° (solid line), the

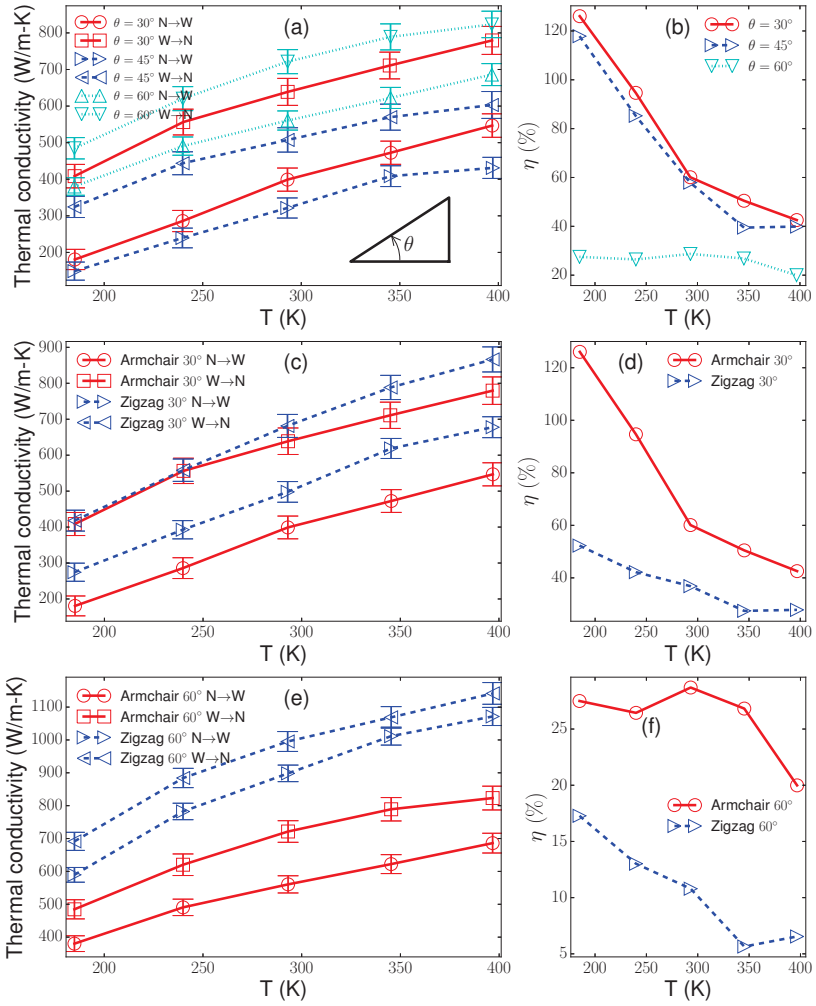

Fig. 2.9. Thermal conductivity of triangular GNRs: (a) dependence of thermal conductivity on vertex angles for triangular GNRs with armchair bottom edges and effect of edge chirality of the bottom edge on thermal conductivity at vertex angle of (b) 30° and (c) 60°. The corresponding thermal rectification factors versus temperature for the various structures simulated are shown in (b), (d) and (f).

GNR with vertex angle of 45° (60°) has lower (higher) thermal conductivity in both directions, but both GNRs with vertex angles of 45° and 60° exhibit less thermal rectification (see Fig. 2.9b). For the GNR with vertex angle of 45°, its hypotenuse has irregular edge and the phonon scattering at this edge likely decreases the thermal conductivity as well as the thermal rectification. In triangular GNRs with armchair bottom edge, only the GNRs with vertex angle of 30° and 60° do not have irregular hypotenuse edges. At large vertex angles, the triangular GNRs approach the symmetric rectangular GNRs, which has zero thermal rectification. Since both irregular edge scattering and large vertex angles can decrease the thermal rectification, we suggest that a vertex angle of 30° may be optimal for thermal rectification. We have demonstrated in Fig. 2.5 that the symmetric zigzag GNRs have larger thermal conductivity than that of the armchair GNRs. This is also true for asymmetric triangular GNRs. In Fig. 2.9c (Fig. 2.9e), the GNR with armchair bottom edge and vertex angle of 30° (60°) (solid lines) has smaller thermal conductivity than the GNR with zigzag bottom edge and vertex angle of 30° (60°) (dashed lines), but the former has larger thermal rectification. Overall, among various triangular GNRs we investigated, the one with armchair bottom edge and vertex angle of 30° has the largest thermal rectification.

In reality, GNRs inevitably have defects. We have studied the effect on the thermal conductivity of GNRs due to two types of defects: circular vacancies and edge roughness. In Fig. 2.10a, we show the thermal conductivity of GNR with single (dotted line) and double (dashed line) circular vacancies and rough edges (dash-dot line). Here a single circular vacancy is created by removing all six carbon atoms of a hexagon. Compared to the perfect rectangular GNR (solid line), the thermal conductivity decreases after introducing circular vacancies. The edge roughness of the symmetric GNR can also decrease its thermal conductivity. This is in qualitative agreement with recent theoretical prediction by Nika et al. that the thermal con-

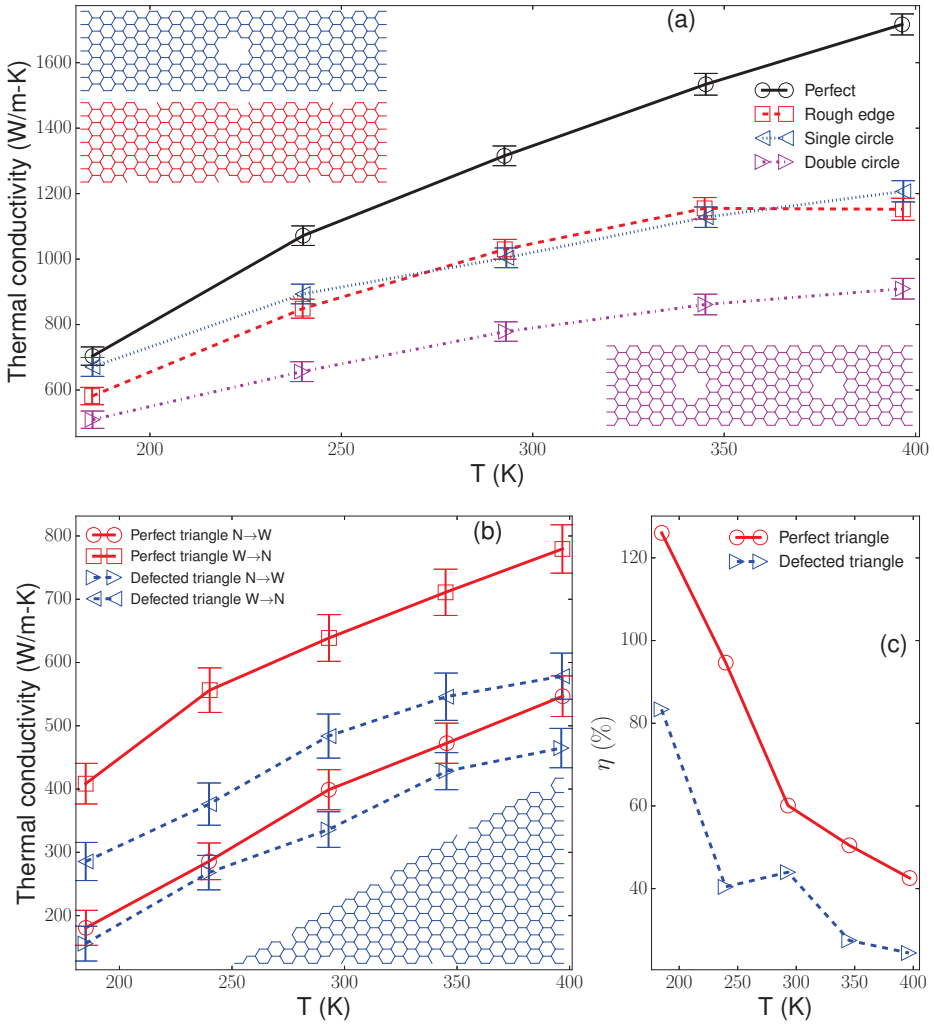

Fig. 2.10. Effect of various defects on thermal conductivity of GNRs: (a) single and double vacancies and edge roughness for a symmetric rectangular GNR, (b) edge roughness for an asymmetric triangular GNR and (c) its thermal rectification factor. The insets are the structure of GNRs with defects.

ductivity of graphene flakes depends on the edge roughness and defect concentration, especially for small flakes [71]. The effect of edge roughness in the triangular GNR

is also studied. We remove six atoms at the bottom edge and the hypotenuse of the triangular GNR of Fig. 2.10b. The solid lines in Fig. 2.10b represent the thermal conductivity of the perfect triangular GNR with armchair bottom edge and vertex angle of 30°. The edge roughness decreases both the thermal conductivity and the thermal rectification, which is nonetheless still as large as 80% at $T \sim 180$ K.

Our collaborators offer a microscopic explanation of TR based on phonon lateral confinement [27]. Three mechanisms that can lead to TR are uncovered: phonon spectra overlap, inseparable dependence of thermal conductivity on temperature and location and phonon edge localization. The overlap between the phonon spectra at two ends of asymmetrical GNRs can be significantly changed when the temperatures at two ends are swapped. The larger the overlap is, the smaller the mismatch is. It is found that phonon spectra mismatch of the out-of-plane motion is dominant over that of the in-plane motion. Since the width of the GNR is different at different locations, this will induce the position dependent phonon mean free path, phonon density of states and thus the local thermal conductivity has complicated dependence on position and temperature. Quite often this dependence is inseparable. It has been demonstrated that such inseparable dependence can lead to TR in bulk materials [72, 73].

TR is also found by our collaborators using MD simulations in other graphene-based systems, e.g., graphene-nanotube hybrids [28,74] and inhomogeneously strained symmetrical GNRs [75].

At almost the same time as our publication on thermal rectification in asymmetrical GNRs, similar results are obtained by other groups in asymmetrical GNR structures [76, 77]. Significant thermal rectification has been obtained.

2.3 Isotopic engineering of thermal conductivity of graphene nanoribbons

Previous studies [78, 79] and ours show that the thermal transport in GNRs depends on the edge chirality of GNRs. In realistic graphene samples, the edges are often passivated [80–82] and the isotope composition can be controlled [83]. We thus study the effect of the edge H-passivation and various isotope distributions on the thermal transport in GNRs. We find that the thermal conductivity can be reduced by the edge H-passivation and tuned by the isotope distributions.

In this section, all the calculated thermal conductivities are normalized by κ_0 (~ 670 W/m-K) which is the thermal conductivity calculated at 100 K for the pure ^{12}C GNR with armchair edge and without H-passivation (shown in Fig. 2.1a). Although the specific value of κ_0 depends on the GNR size, the choice of thermostat and boundary condition in MD simulation, we have checked that our conclusions and the qualitative behavior of κ discussed below do not.

We study the temperature dependent thermal conductivity of H-passivated GNRs. All of our simulation results indicate that the thermal conductivity increases with temperature. This behavior is consistent with our previous work. Fig. 2.11 shows the armchair and zigzag GNRs with top and bottom edges H-passivated. As shown in Fig. 2.12, the edge H-passivation significantly reduces the thermal conductivity, compared to the non-passivated GNRs. A recent study [84] using equilibrium MD has obtained similar conclusions. We note that the error bars (related to molecular dynamics fluctuations) for H-passivated GNRs are considerably larger than that for the non-passivated GNRs, probably due to the much smaller mass of hydrogen atoms.

We also study the effect of the mixture of carbon isotopes ^{12}C and ^{13}C on the thermal conductivity of GNRs. Here, we demonstrate the results in the case of armchair GNRs (qualitatively similar results are obtained for zigzag GNRs). The concentration

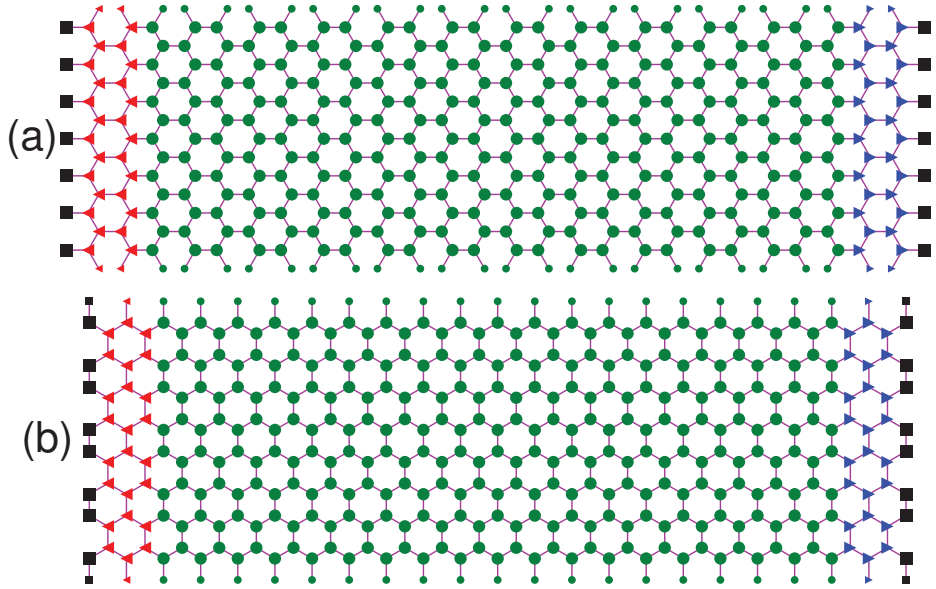

Fig. 2.11. Structure of hydrogen-passivated armchair (a) and zigzag (b) GNRs. The hydrogen atoms are denoted by smaller symbols while the ^{12}C atoms are denoted by larger ones.

of ^{13}C is $N_{13}/(N_{12}+N_{13})$, where $N_{12/13}$ is the number of $^{12/13}$C atoms. The thermal conductivity is seen to be reduced by introducing ^{13}C, and the thermal conductivity of pure ^{13}C GNRs is lower than that of pure ^{12}C GNRs because ^{13}C atoms have larger mass and thus give lower phonon frequency [85]. The inset of Fig. 2.13a shows a typical random isotope distribution. Another isotope distribution pattern we study is the isotopic superlattice structure shown in the inset of Fig. 2.13b. Here, the whole GNR is composed of four slices with equal length and the same isotope composition. Within each slice (such as the shaded area in the inset of Fig. 2.13b), the number (L) of vertical ^{13}C atomic chains with zigzag shape (e.g., $L = 4$ for the inset of Fig. 2.13b) can vary from 0 to 7. $L = 0$ ($L = 7$) corresponds to the pure ^{12}C (^{13}C) GNR. The

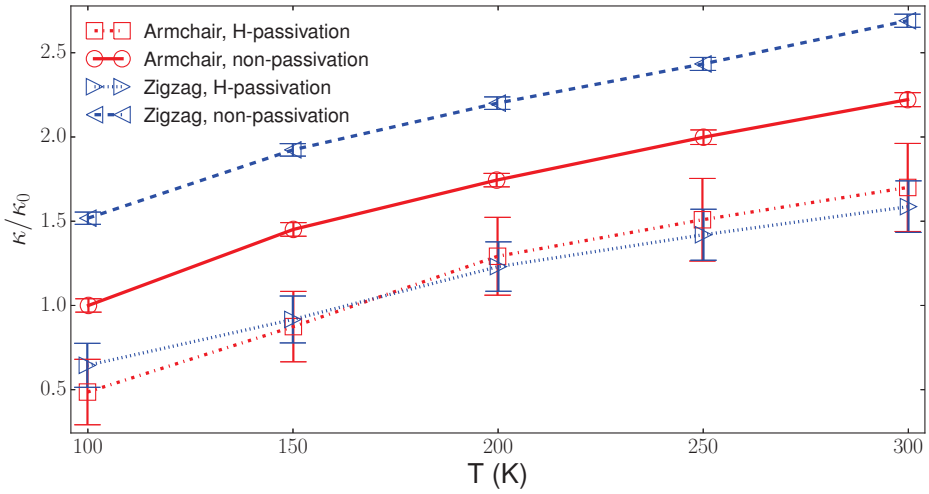

Fig. 2.12. Temperature dependent thermal conductivity of GNRs with and without edge H-passivation.

temperature dependent thermal conductivity in Fig. 2.13 shows that the isotope effect becomes more evident at higher temperatures. We show the thermal conductivity as a function of the concentration of ^{13}C calculated at the temperature of 500 K in Fig. 2.14. In the case of the random distributions, the calculated thermal conductivity is an average of 10 different distributions with the same isotope concentration. For random isotope distributions, the isotope concentration dependent thermal conductivity (dashed line in Fig. 2.14) shows a pan shape and is relatively flat in the concentration range of $\sim 20 - 90\%$. The thermal conductivity is reduced by $\sim 10\%$ around the isotope concentration of $\sim 50\%$. The error bars are determined from the deviations of the thermal conductivities for the 10 different distributions from their average value. In contrast, the conical shape of the solid line in Fig. 2.14 shows much stronger dependence of the thermal conductivity on the isotope concentration for the superlattice structures, with $\sim 30\%$ reduction of the thermal conductivity at $\sim 50\%$

of the isotope concentration. We have also obtained similar results using velocity exchange MD [86] in the LAMMPS package [87].

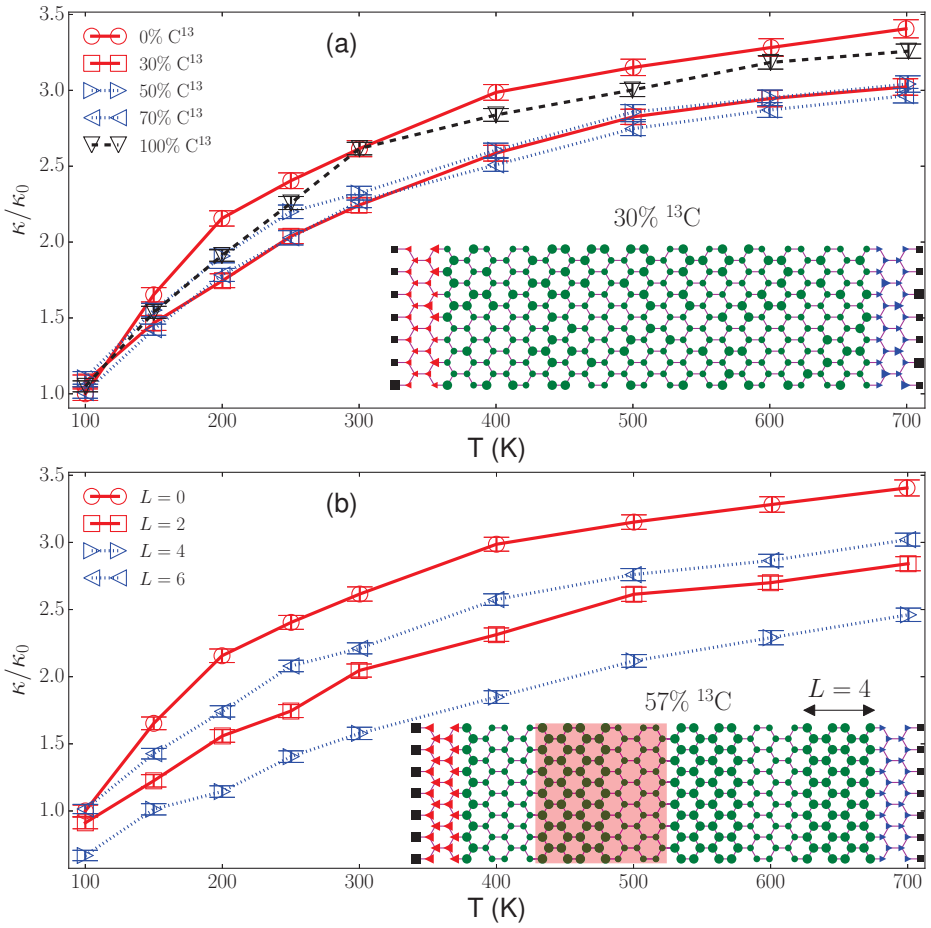

Fig. 2.13. Temperature dependent thermal conductivity of GNRs with ^{13}C isotopes distributed (a) randomly and (b) in a superlattice structure. The insets show the corresponding typical structures of GNRs with the same meaning of symbols as that in Fig. 2.1. The larger (smaller) symbols denote ^{13}C (^{12}C) atoms.

We have demonstrated that the thermal conductivity depends on the edge chirality in the absence of H-passivation, i.e., the thermal conductivity of the zigzag GNR is larger than that of the armchair GNR. However, as shown in Fig. 2.12, their thermal conductivities become close to each other within the MD error bars after the H-passivation, suggesting that phonon scattering with the hydrogenated edges dominates over the contribution from the chirality effect. We have suggested that the smaller thermal conductivity of armchair GNRs is due to the stronger phonon scattering at the armchair edges. It is interesting to note that the H-passivated zigzag GNR in Fig. 2.11b resembles the armchair GNRs at the edges. We suggest that the thermal transport in small GNRs (several nanometers in size in our study) is strongly affected by the edge configuration.

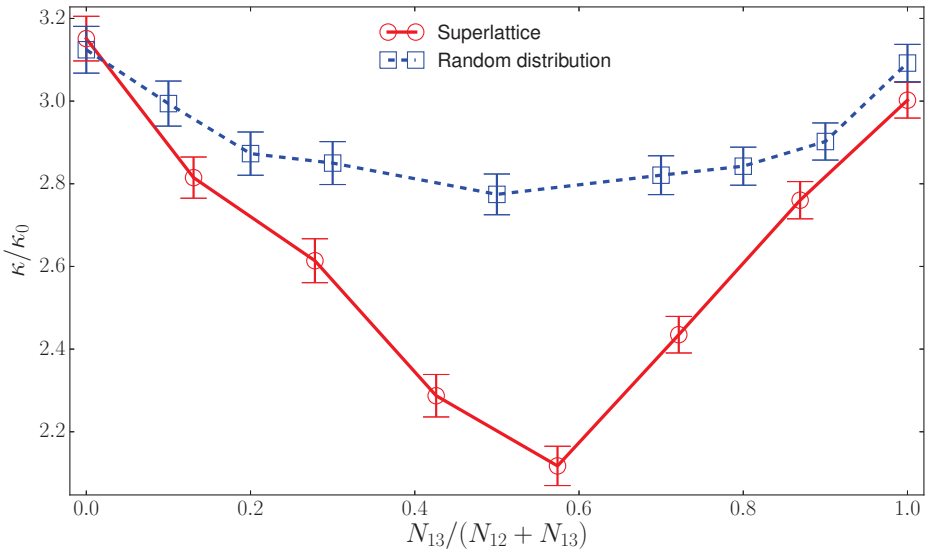

Fig. 2.14. Thermal conductivity as a function of the ^{13}C concentration for superlattice (solid line) and random (dashed line) isotope distributions.

Recently, it has been experimentally demonstrated that different carbon isotopes can be controllably introduced in graphene, such as ^{13}C, in the chemical vapor deposition growth of graphene on metals. Both random and segregation (by domains of different isotopes) distributions have been observed [83]. This opens possibilities of engineering the thermal properties of graphene by isotope distributions. The isotope effect on the thermal transport has been studied in several nanomaterials, such as carbon nanotubes [88, 89], boron nitride nanotubes [90, 91], and silicon nanowires [92]. The pan shape of the dashed line in Fig. 2.14 is consistent with the the reduction of the thermal conductivity in the "alloy limit" [93]. A similar pan shape is found in GNRs [94] and SiGe nanowires [95] by tuning the composition. By keeping the isotope concentration a constant of $\sim 50\%$, it has been shown [96] that the thermal conductivity as a function of the slice length (which is kept constant in our simulations) gives similar conical shaped dependence as we see in Fig. 2.14 for the superlattice structures.

2.4 Nonlinear thermal transport in graphene nanoribbons

We notice that little attention has been paid to nonlinear thermal transport in GNRs, though these nonlinear effects have been explored in ideal atomic chains [56, 97–100], molecular junctions [101] and quantum dots [102]. In GNRs, we demonstrate negative differential thermal conductance (NDTC), an unusual thermal transport phenomenon where the heat current across a thermal conductor decreases when the temperature bias increases. It is shown that NDTC exists in many non-linear one-dimensional (1D) systems [25, 56, 97–100, 103–107] and vacuum gaps [108]. Many corresponding mechanisms such as nonlinear interactions [101], molecular anharmonicity [109], competition mechanisms [97] and others [108] are proposed to explain the

existence of NDTC. Analogous to the electronic counterpart [110], NDTC is a useful ingredient in developing GNR-based thermal management and signal manipulation devices, such as the thermal amplifiers [56] and thermal logic gates [57].

First, we study the nonlinear thermal transport in a rectangular GNR shown in Fig. 2.1a (we have obtained qualitatively similar conclusions for GNRs with zigzag edges). Since the GNR is symmetrical, we only consider $T_{\rm L} \leq T_{\rm R}$ and define the temperature difference $\Delta T \equiv T_{\rm R} - T_{\rm L}$. The temperature $T_{\rm R}$ is kept as a constant. As we can see from both curves in Fig. 2.17, for small temperature difference (e.g., $\Delta T < 60$ K for $T_{\rm R} = 300$ K and $\Delta T < 150$ K for $T_{\rm R} = 600$ K), the thermal current increases approximately linearly as ΔT increases, as expected from Fourier's law. Interestingly, for some range of higher ΔT, the thermal current decreases as ΔT increases (the dashed boxes in Fig. 2.17), indicating the onset of negative differential thermal conductance (NDTC). It is a reasonable approximation to consider thermal current as proportional to the product of thermal conductivity κ of the GNR and ΔT. We has shown that κ increases with the average temperature $\overline{T} \equiv (T_{\rm L} + T_{\rm R})/2 = T_{\rm R} - \Delta T/2$ (see Fig. 2.5). We have plotted $\overline{T}$ (labeled at the right vertical axis and indicated by right-pointing arrows for Fig. 2.15, Fig. 2.16, Fig. 2.17) and in the subplot of Fig. 2.18 as a function of ΔT in all figures. Since $\overline{T}$ decreases with ΔT, κ decreases with increasing ΔT. The resulting trend of the thermal current as a function of ΔT is thus a competition between decreasing κ and increasing ΔT. In the ΔT range displaying NDTC, the decrease of κ with ΔT dominates. We have found that there is no NDTC (shown in Fig. 2.18) if $T_{\rm L}$ is larger than the constant $T_{\rm R}$, i.e., if $\overline{T}$ increases with ΔT (thus without the above competition). Note that for large ΔT beyond linear response, strictly speaking thermal conductivity is not well defined. Thus, in the above explanation, κ is considered to be an effective, average thermal

conductivity. Similar arguments have been applied in analyzing thermal transport in 1D atomic chains [100].

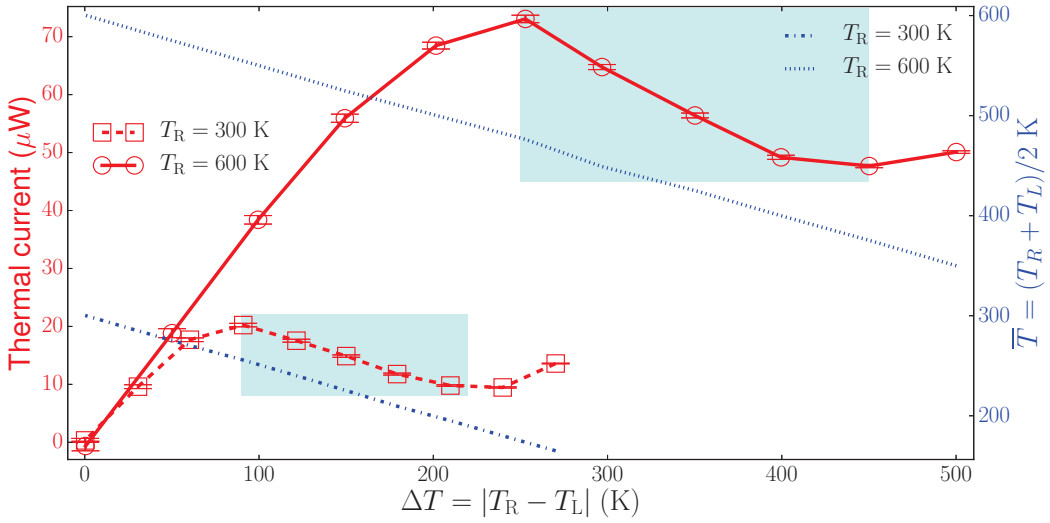

Fig. 2.15. Thermal current vs. temperature difference ΔT. The dashed boxes exhibit NDTC for an armchair GNR ($\sim$ 1.5 nm $\times$ 6 nm).

Second, we study the length dependence of NDTC in GNRs. For all three GNRs of different lengths in Fig. 2.16, $T_R = 300$ K while T_L is varied from T_R to 30 K. As the GNR length is increased, the ΔT value for the onset of NDTC increases and the ΔT range where NDTC exists shrinks. We thus suggest that NDTC will eventually disappear if the length of GNR exceeds some critical value. We have verified this using LAMMPS package [87] and velocity scaling [111] MD, and found no NDTC in a 50 nm long GNR with similar structure as that studied in Fig. 2.15.

Besides these nonlinear effects in symmetrical GNRs, we also explore the possibility of NDTC in an asymmetrical triangular GNR, shown in Fig. 2.1c. Our previous study has pointed out that thermal rectification exists in this asymmetrical GNR.

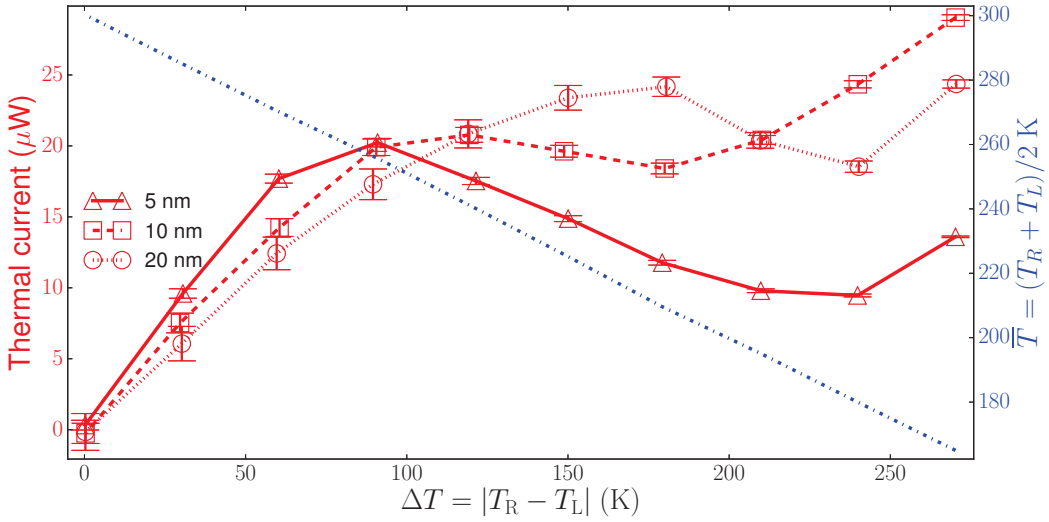

Fig. 2.16. Thermal current vs. temperature difference ΔT in GNRs with the similar structure as the symmetrical GNR in Fig. 2.15, except for different lengths. In all these plots, $T_{\rm R} = 300$ K and $T_{\rm L}$ is varied from 300 K to 30 K.

As we see from Fig. 2.17, here the nonlinear thermal transport is also direction-dependent. NDTC appears when the temperature of the narrower end is held at $T_{\rm L} = 300$ K and the temperature $T_{\rm R}$ of the wider end is varied from 300 K to 30 K (solid line in Fig. 2.17). However, there is no NDTC when the values of $T_{\rm L}$ and $T_{\rm R}$ are interchanged (dashed line in Fig. 2.17). This provides another possibility to control the nonlinear thermal transport and NDTC in GNRs by engineering the shape of GNRs.

In general, the way to tune the thermal current in the two-terminal thermal devices is very different from that in any two-terminal electronic devices. In the latter case, only the voltage difference matters. However, in thermal devices, the average temperature $\overline{T}$ is as important as the temperature difference ΔT in controlling the

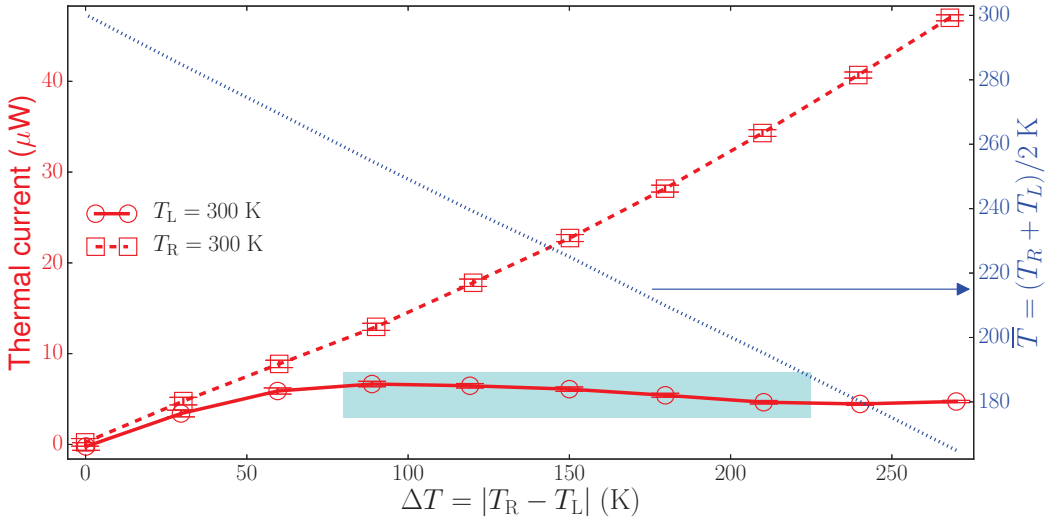

Fig. 2.17. Thermal current vs. temperature difference ΔT in triangular GNR shown in Fig. 2.1. The dashed box exhibits NDTC.

thermal current. For example, consider $\overline{T} = \alpha\Delta T + T_0$ with constants α and T_0, and we have $T_{\rm L} = (\alpha - \frac{1}{2})\Delta T + T_0$ and $T_{\rm R} = (\alpha + \frac{1}{2})\Delta T + T_0$. The thermal currents and average temperature $\overline{T}$ as a function of ΔT are plotted in Fig. 2.18 for the rectangular GNR shown in the inset in Fig. 2.1a, where $T_0 = 300$ K and α is tuned from -0.5 to 0.5 (indicated by the dashed curved arrow in Fig. 2.18). The solid curve in Fig. 2.15 corresponds to $\alpha = -0.5$. For small temperature difference in the linear response regime, the slope of thermal current vs. ΔT is independent of α. In the nonlinear response regime (large ΔT), the system transitions from a regime with NDTC to a regime without NDTC when α is tuned from negative to positive values. We can see a strong correlation between the the trend of the thermal current and that of the average temperature for different values of α in the range of ΔT from 100 K to 250 K where NDTC occurs for negative α. For negative α, since $\overline{T}$ decreases with

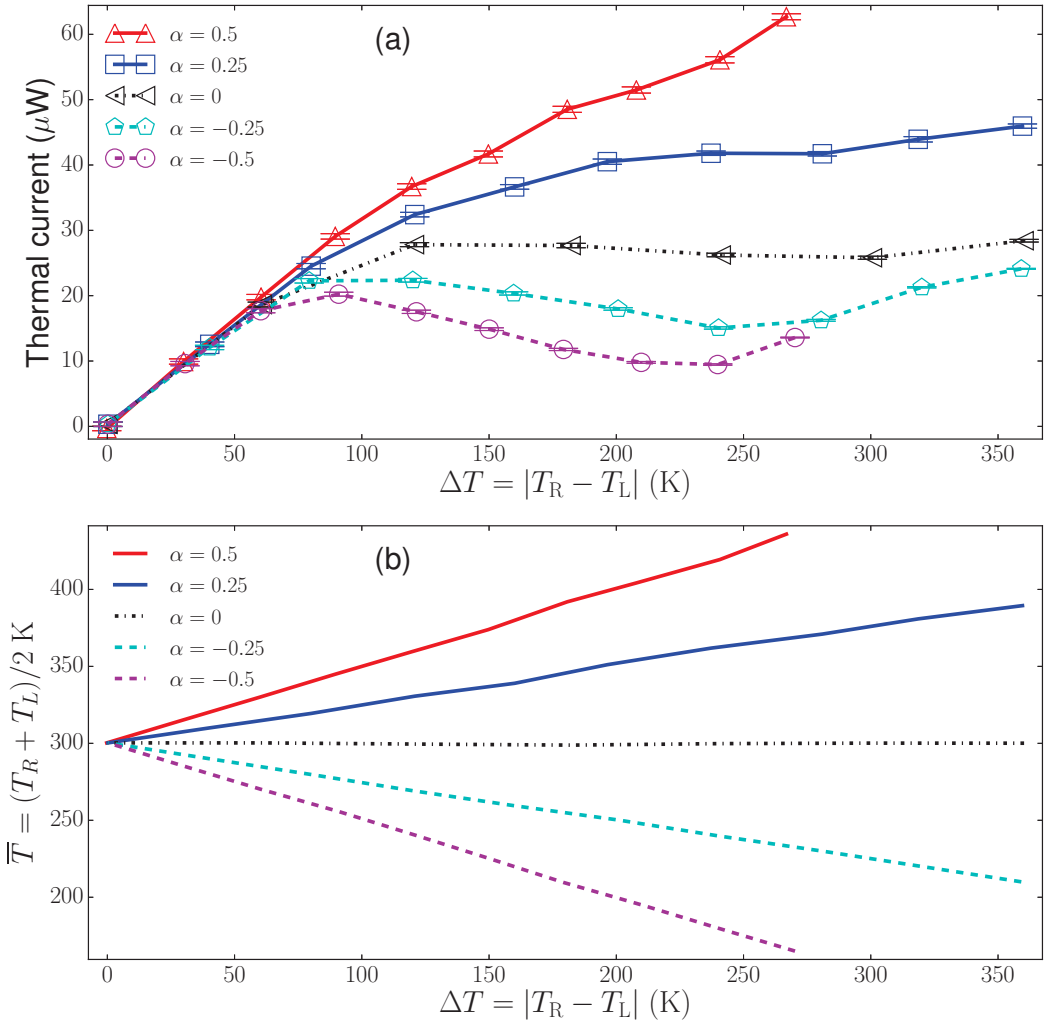

Fig. 2.18. Thermal current vs. temperature difference ΔT for different values of α for the symmetrical GNR. Note that $\alpha = 0.5$ (-0.5) corresponds to $T_{\rm L(R)}$ fixed at 300 K while $T_{\rm R(L)}$ is varied.

ΔT, the effective κ decreases with ΔT, and the occurrence of NDTC can be similarly explained as that for Fig. 2.15.

There are two independent parameters to control the thermal transport in two-terminal devices, either $(T_{\mathrm{L}}, T_{\mathrm{R}})$ or $(\Delta T, \overline{T})$. Two-terminal thermal devices are actually analogous to three-terminal electronic devices. In the language of electronic transport of field effect transistors (FETs), ΔT plays the role of the drain-source voltage difference in FETs, while α plays the role of the gate voltage. Fig. 2.18 shows the ability to realize the FET-like behaviour in GNRs.

2.5 Existence of negative differential thermal resistance in 1D diffusive transport

Many studies [100, 104] and ours suggest that NDTC may vanish as the system length becomes large (approaching diffusive thermal transport). Here, we study the conditions for the existence of NDTC for 1D diffusive thermal transport described by Fourier's law. We prove that NDTC cannot exist when the temperature at one end is fixed. However, we show that NDTC is still possible if a thermal contact resistance (TCR) is introduced [112].

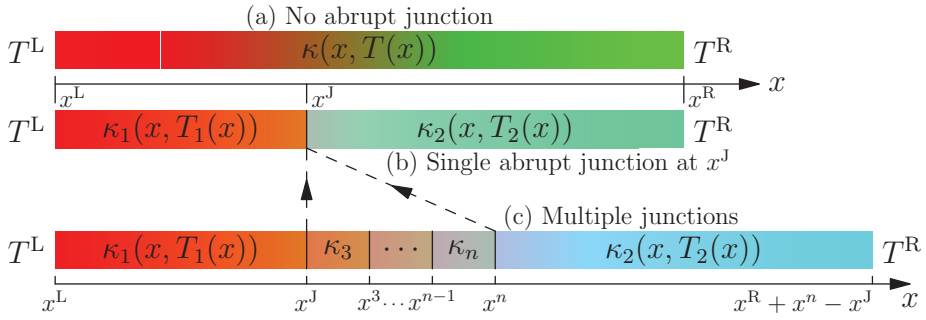

Fig. 2.19. Schematics of 1D systems without any junctions (a) and with a single junction (b) and multiple junctions (c). The junctions are indicated by vertical black lines.

When considering a general 1D system in the diffusive regime, the system can be described by its thermal conductivity $\kappa(x, T)$ as a function of the coordinate x and the local temperature $T(x)$. Here we have implicitly assumed that the system is in local thermal equilibrium so that the local quantities are well defined from statistical point of view. More precisely, we can imagine that the system is divided into many small segments, each of them big enough to contain enough number of atoms so that the the thermodynamic quantities (e.g., temperature and heat current) can be well defined over the segment, yet small enough on the macroscopic scale so that for slow variation of the system in space we can use continuous description $T(x)$ and $\kappa(x, T(x))$. This phenomenological description is valid as long as the mean free path (MFP) of heat carriers within each segment is much smaller than the size of that segment.

The position dependence of the thermal conductivity $\kappa(x, T(x))$ is explicitly expressed rather than implicitly by the position dependence of temperature, since the systems we consider might have their compositions (e.g., mass density graded systems) and/or structures (e.g., strain) continuously changing in space and at different positions the behaviors of the temperature dependence of thermal conductivities are different.

We believe the phenomenological description of thermal transport in a large class of 1D systems in diffusive regime using $\kappa(x, T(x))$ is general and enough to consider the existence of NDTC, without running into any microscopic details of these 1D systems. This approach provides analytic results regarding to the existence of NDTC, and it is instructive in system design to pursue the applications of NDTC effect.

The heat current can be calculated from the Fourier's law:

$$q = -\kappa(x, T)\frac{\mathrm{d}T}{\mathrm{d}x}. \tag{2.6}$$

For thermal transport without heat sources or sinks, the heat current is conserved and the steady state thermal transport equation reads

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\kappa(x,T)\frac{\mathrm{d}T}{\mathrm{d}x}\right)=0. \tag{2.7}$$

Given appropriate boundary conditions, we can in principle calculate the heat current and the temperature profile of the system, and furthermore the differential thermal conductance (DTC).

2.5.1 Systems without abrupt junctions

We first consider a finite 1D system which lies in the coordinate range $[x^{\mathrm{L}}, x^{\mathrm{R}}]$ without any abrupt thermal contact junctions. Systems without abrupt junctions are shown in Fig. 2.19a. In this case, if $T^{\mathrm{R(L)}}$ is fixed (without loss of generality, we can assume $T^{\mathrm{L}} > T^{\mathrm{R}}$), we will show that there is no NDTC and the heat current q will increase as the temperature bias increases by increasing (lowering) $T^{\mathrm{L(R)}}$. NDTC could still exist when the temperatures at two ends vary simultaneously. if the temperatures at both ends vary simultaneously and depend on a parameter u (i.e., $T^{\mathrm{L(R)}}(u)$), the DTC is $G=\frac{G^{\mathrm{L}}dT^{\mathrm{L}}/du-G^{\mathrm{R}}dT^{\mathrm{R}}/du}{dT^{\mathrm{L}}/du-dT^{\mathrm{R}}/du}$. It is possible to have the numerator of G to be negative while its denominator is positive, if the form of $T^{\mathrm{L(R)}}(u)$ is designed carefully. However, we limit our study in the cases that the temperature at one end is fixed. Once the temperature at two ends of the system are given, i.e.,

$$T(x^{\mathrm{L(R)}})=T^{\mathrm{L(R)}}, \tag{2.8}$$

the temperature profile $T(x)$ can be calculated from Eq. (2.7) and boundary conditions Eq. (2.8), and the resulting heat current q (independent of x) flowing in the system can be computed from Eq. (2.6).

By applying an infinitesimal variation $\delta T^{\mathrm{L(R)}}$ of the boundary temperature at one end, i.e., $T^{\mathrm{L(R)}}$ is varied to $T^{\mathrm{L(R)}} + \delta T^{\mathrm{L(R)}}$ while $T^{\mathrm{R(L)}}$ is fixed, the resulting temperature profile is varied to $T(x) + \delta T(x)$. We must have

$$\delta T(x^{\mathrm{L(R)}}) = \delta T^{\mathrm{L(R)}} \text{ while } \delta T(x^{\mathrm{R(L)}}) = 0. \tag{2.9}$$

This temperature profile variation $\delta T(x)$ can induce a variation δq of the heat current. We define the DTC as

$$G \equiv \frac{\delta q}{\delta(T^{\mathrm{L}} - T^{\mathrm{R}})}, \tag{2.10}$$

or specifically

$$G = G^{\mathrm{L}} = \frac{\delta q}{\delta T^{\mathrm{L}}}, \text{ when } T^{\mathrm{R}} \text{ is fixed; or } G = G^{\mathrm{R}} = -\frac{\delta q}{\delta T^{\mathrm{R}}}, \text{ when } T^{\mathrm{L}} \text{ is fixed.} \tag{2.11}$$

Qualitatively, we first point out that the non-existence of NDTC here is a direct consequence of the uniqueness of the solution of Eq. (2.7), as we graphically demonstrate in Fig. 2.20a. Take the case that T^{R} is fixed as an example. If the temperature T^{L} is increased to T'^{L} ($> T^{\mathrm{L}}$), the temperature profile $T(x)$ (black line) with $T(x^{\mathrm{L(R)}}) = T^{\mathrm{L(R)}}$ and heat current q are changed to $T'(x)$ (red line) with $T'(x^{\mathrm{L}}) = T'^{\mathrm{L}}$ and $T'(x^{\mathrm{R}}) = T^{\mathrm{R}}$ and heat current q'. First of all, we must have $q' \neq q$. Otherwise, the first order differential equation $\frac{\mathrm{d}T}{\mathrm{d}x} = -\frac{q}{\kappa(x,T)}$ about T with initial condition $T(x^{\mathrm{R}}) = T^{\mathrm{R}}$ would have non-unique solutions ($T(x)$ and $T'(x)$), which is not allowed. Secondly, q' cannot be smaller than q (proportional to the slope of $T(x)$ at x^{R}), because otherwise there will be an intersection of $T'(x)$ (represented by the dashed line) and $T(x)$ at some $x^{\mathrm{I}} < x^{\mathrm{R}}$. We then must have $T'(x) = T(x)$ in the coordinate range $[x^{\mathrm{I}}, x^{\mathrm{R}}]$ due to the uniqueness of the solution to Eq. (2.7), and thus $q' = q$ (contradiction). Therefore, we must have $q' > q$, i.e., the heat current monotonically increases with temperature T^{L} when T^{R} is fixed and there is no NDTC. Similar arguments apply to the case when T^{L} is fixed.

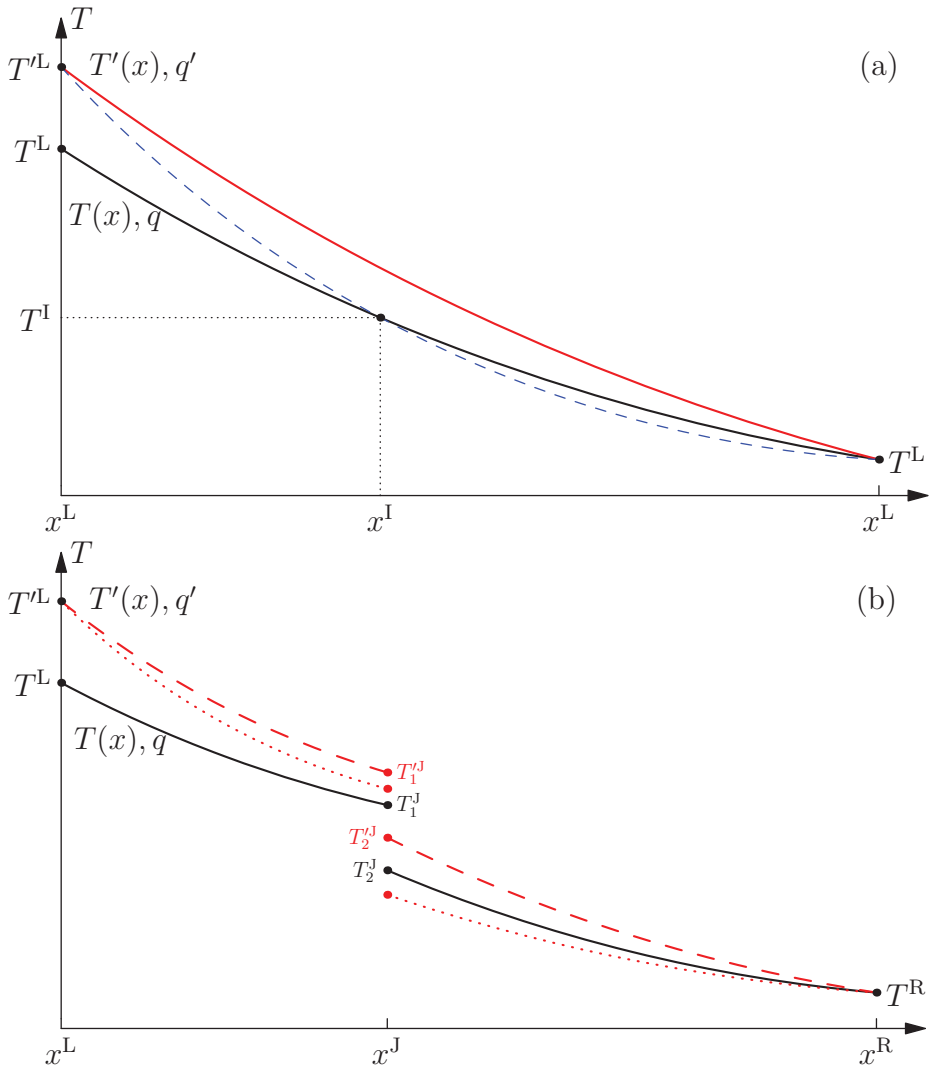

Fig. 2.20. A schematic example of temperature profiles of systems (a) without any junctions and (b) with a single junction at x^{J} when the temperature at x^{R} is fixed and the temperature at x^{L} is increased from T^{L} to T'^{L}. The dotted lines in (b) would give rise to NDTC.

To calculate the DTC, we start from the variation of Eq. (2.6):

$$\delta q = -\kappa(x,T)\frac{\mathrm{d}}{\mathrm{d}x}\delta T - \frac{\partial\kappa}{\partial T}\frac{\mathrm{d}T}{\mathrm{d}x}\delta T. \tag{2.12}$$

We define

$$U^{\mathrm{L(R)}}(x) \equiv \frac{\delta T(x)}{\delta T^{\mathrm{L(R)}}} \tag{2.13}$$

which according to Eq. (2.12) satisfies

$$\frac{\mathrm{d}}{\mathrm{d}x}U^{\mathrm{L(R)}} + \frac{1}{\kappa(x,T)}\frac{\partial\kappa}{\partial T}\frac{\mathrm{d}T}{\mathrm{d}x}U^{\mathrm{L(R)}} + \frac{\eta^{\mathrm{L(R)}}G^{\mathrm{L(R)}}}{\kappa(x,T)} = 0. \tag{2.14}$$

where $\eta^{\mathrm{L}} = 1$ and $\eta^{\mathrm{R}} = -1$. From Eqs. (2.9) and (2.13), the boundary conditions for Eq. (2.14) are

$$U^{\mathrm{L}}(x^{\mathrm{L}}) = 1, U^{\mathrm{L}}(x^{\mathrm{R}}) = U^{\mathrm{R}}(x^{\mathrm{L}}) = 0, U^{\mathrm{R}}(x^{\mathrm{R}}) = 1. \tag{2.15}$$

Eq. (2.14) is an inhomogeneous linear ordinary differential equation, and the coefficients $\frac{1}{\kappa(x,T)}\frac{\partial\kappa}{\partial T}\frac{\mathrm{d}T}{\mathrm{d}x}$ and $\frac{\eta^{\mathrm{L(R)}}G^{\mathrm{L(R)}}}{\kappa(x,T)}$ are functions of x only, since $T(x)$ is already formally solved from Eq. (2.7) and boundary conditions Eq. (2.8). The solution to Eq. (2.14) is

$$U^{\mathrm{L(R)}}(x) = \frac{U^{\mathrm{L(R)}}(x^{\mathrm{L}})}{F(x)} - \frac{\eta^{\mathrm{L(R)}}G^{\mathrm{L(R)}}}{F(x)}\int_{x^{\mathrm{L}}}^{x}\frac{F(t)}{\kappa(t,T(t))}\mathrm{d}t,iman \tag{2.16}$$

where

$$F(x) \equiv \exp\left(\int_{x^{\mathrm{L}}}^{x}\frac{1}{\kappa(t,T(t))}\frac{\partial\kappa}{\partial T}\frac{\mathrm{d}T}{\mathrm{d}t}\mathrm{d}t\right). \tag{2.17}$$

By evaluating Eq. (2.16) at $x = x^{\mathrm{R}}$, we obtain

$$G^{\mathrm{L(R)}} = J\eta^{\mathrm{L(R)}}\left[U^{\mathrm{L(R)}}(x^{\mathrm{L}}) - F(x^{\mathrm{R}})U^{\mathrm{L(R)}}(x^{\mathrm{R}})\right], \tag{2.18}$$

where

$$J \equiv \left[\int_{x^{\mathrm{L}}}^{x^{\mathrm{R}}}\frac{F(t)}{\kappa(t,T(t))}\mathrm{d}t\right]^{-1}. \tag{2.19}$$

From Eq. (2.15) and Eq. (2.18), we have

$$G^{\mathrm{L}} = J, \;\; G^{\mathrm{R}} = F(x^{\mathrm{R}})J. \tag{2.20}$$

Since $F(x)$ and κ are positive, we have $G^{\mathrm{L}} > 0$ and $G^{\mathrm{R}} > 0$. Namely, if $T^{\mathrm{R(L)}}$ is fixed, the heat current q will increase as the temperature bias increases by increasing (lowering) $T^{\mathrm{L(R)}}$. Therefore, there is no NDTC for systems without abrupt junctions. A geometrical view of point is provided here to elaborate this non-existence of NDTC.

In general, the heat current $q = q(T^{\mathrm{R}}, T^{\mathrm{L}})$ is a function of T^{R} and T^{L} and its variation is

$$\delta q = \frac{\partial q}{\partial T^{\mathrm{L}}}\delta T^{\mathrm{L}} + \frac{\partial q}{\partial T^{\mathrm{R}}}\delta T^{\mathrm{R}} = G^{\mathrm{L}}\delta T^{\mathrm{L}} - G^{\mathrm{R}}\delta T^{\mathrm{R}}. \tag{2.21}$$

Let's consider the contour lines of constant q on the quarter plane of $(T^{\mathrm{R}}, T^{\mathrm{L}}) \in [0, \infty) \times [0, \infty)$. These contour lines are described by

$$G^{\mathrm{L}}\mathrm{d}T^{\mathrm{L}} = G^{\mathrm{R}}\mathrm{d}T^{\mathrm{R}} \tag{2.22}$$

by assigning $\delta q = 0$. It is clear that $\frac{\mathrm{d}T^{\mathrm{L}}}{\mathrm{d}T^{\mathrm{R}}} = F(x^{\mathrm{R}}) > 0$, thus every contour line has a positive slope and we can conclude that each contour line can intersect with any line of constant T^{L} or T^{R} at most once. In other words, Eq. (2.22) poses a global topological constraint on the contour lines: the heat current at the intersection points of any constant $T^{\mathrm{R(L)}}$ line with all contour lines has the same ordering as that at the intersection points of the $T^{\mathrm{R(L)}}$ axis with all contour lines. This ordering property of the contour lines is a consequence of the mathematical form of Eq. (2.7), regardless of the choice of $\kappa(x, T)$ (e.g., even for non-physical systems with $\kappa < 0$). Furthermore, from Eq. (2.20) we already know that $q(0, T^{\mathrm{L}})$ ($q(T^{\mathrm{R}}, 0)$) monotonically increases (decreases) with $T^{\mathrm{L(R)}}$. The topological constraint and the monotonicity of $q(0, T^{\mathrm{L}})$ and $q(T^{\mathrm{R}}, 0)$ ensure that the heat current $q(T^{\mathrm{R}}, T^{\mathrm{L}})$ is a monotonic function of $T^{\mathrm{L(R)}}$ when $T^{\mathrm{R(L)}}$ is fixed, which implies that NDTC is impossible along any lines

of constant T^{L} or T^{R}. The non-existence of NDTC is a physical consequence of Eq. (2.7). Namely, it is because of the positivity of the thermal conductivity (i.e., heat always flows from hotter to colder regions) of any physical system obeying the second law of thermodynamics.

It is worth to mention that the above topological ordering of contour lines is preserved even when κ depends on the temperature gradient $\frac{\mathrm{d}T}{\mathrm{d}x}$. This dependence is physically possible. For example, in the large heat current regime, the local thermal conductivity may depend on the local heat current flow. If $\kappa = \kappa(x, T, \frac{\mathrm{d}T}{\mathrm{d}x})$ instead of $\kappa = \kappa(x, T)$, it is possible to have NDTC exist along lines of constant T_{L} or T_{R}. Once NDTC exists along some line of constant T_{L} or T_{R}, it will exist along every line of constant T_{L} or T_{R}.

2.5.2 Systems with abrupt junctions

Systems with a single abrupt junction are shown in Fig. 2.19b. Let's consider the situation that there is a single abrupt junction at position $x^{\mathrm{J}} \in (x^{\mathrm{L}}, x^{\mathrm{R}})$. The system can be considered as two subsystems without any abrupt junctions coupled together at x^{J}. We suppose that the subsystem in $[x^{\mathrm{L}}, x^{\mathrm{J}}]$ has thermal conductivity $\kappa_1(x, T_1)$ with temperature profile $T_1(x)$ and the subsystem in $[x^{\mathrm{J}}, x^{\mathrm{R}}]$ has thermal conductivity $\kappa_2(x, T_2)$ with temperature profile $T_2(x)$. We denote

$$T_1^{\mathrm{J}} \equiv T_1(x^{\mathrm{J}}),\ T_2^{\mathrm{J}} \equiv T_2(x^{\mathrm{J}}),\ T^{\mathrm{L}} \equiv T_1(x^{\mathrm{L}}),\ T^{\mathrm{R}} \equiv T_2(x^{\mathrm{R}}). \tag{2.23}$$

At the junction, the two subsystems are coupled through a thermal contact resistance (TCR) R^{J}. Usually the TCR is a function of many parameters including the junction temperatures, contact pressure, and other conditions around the junction. We only consider the dependence of junction temperatures by assuming that all the other

conditions remains same during changing the junction temperatures. Any variation of the TCR caused by conditions like contact pressure that are affected by the junction temperature changes should in principle be included in the variations of TCR induced by the change of the junction temperatures.

The possibility of NDTC here can also be interpreted graphically. Again, take the case that T^{R} is fixed as an example. If the temperature T^{L} ($> T^{\mathrm{R}}$) is increased to T'^{L} ($> T^{\mathrm{L}}$), the temperature profile $T(x)$ with $T(x^{\mathrm{L(R)}}) = T^{\mathrm{L(R)}}$, junction temperatures $T^{\mathrm{J}}_{1,2}$ and heat current q are changed to $T'(x)$ with $T'(x^{\mathrm{L}}) = T'^{\mathrm{L}}$ and $T'(x^{\mathrm{R}}) = T^{\mathrm{R}}$, junction temperatures $T'^{\mathrm{J}}_{1,2}$ and heat current q', as illustrated in Fig. 2.20b. Because of the discontinuous jump of temperature at x^{J}, we could have $q' > q$ if $T'^{\mathrm{J}}_2 > T^{\mathrm{J}}_2$ (red dashed line) or $q' < q$ if $T'^{\mathrm{J}}_2 < T^{\mathrm{J}}_2$ (red dotted line) where the latter situation gives rise to NDTC.

The form of the temperature dependence of the TCR is taken to be $R^{\mathrm{J}}(T^{\mathrm{J}}_1, T^{\mathrm{J}}_2)$ and it is justified from the experimental point of view in the following. The properties of the junction are determined by the properties of the materials in the vicinity of the junction, namely, materials in the coordinate range $[x^{\mathrm{J}} - \delta x, x^{\mathrm{J}} + \delta x]$ where δx is an small quantity chosen in such a way that the materials in $[x^{\mathrm{J}} - \delta x, x^{\mathrm{J}}]$ and $[x^{\mathrm{J}}, x^{\mathrm{J}} + \delta x]$ are considered to be effectively homogeneous. On the other hand, the properties of the junction are essentially independent of the materials in $[x^{\mathrm{L}}, x^{\mathrm{J}} - \delta x]$ and $[x^{\mathrm{J}} + \delta x, x^{\mathrm{R}}]$.

To characterize the properties of the junction, we first need the temperature dependence of the thermal conductivities of materials in $[x^{\mathrm{J}} - \delta x, x^{\mathrm{J}}]$ and $[x^{\mathrm{J}}, x^{\mathrm{J}} + \delta x]$ that we assume are already known and are denoted by κ^{L} and κ^{R} respectively. Furthermore, κ^{L} and κ^{R} can be effectively considered to be independent of position since δx is very small. Now we can imagine that two thermal baths are attached to control the temperatures τ^{L} and τ^{R} at $x^{\mathrm{J}} - \delta x$ and $x^{\mathrm{J}} + \delta x$ respectively. The heat current q

flowing across the junction can be determined by measuring the energy transfer rate from one thermal bath to another. The TCR and the heat current satisfy

$$q = \frac{\tau^{\mathrm{L}} - \tau^{\mathrm{R}}}{R^{\mathrm{J}} + \dfrac{\delta x}{\kappa^{\mathrm{L}}} + \dfrac{\delta x}{\kappa^{\mathrm{R}}}}. \tag{2.24}$$

In our thought experiment, δx can be small enough so that the temperature drop is almost entirely on the junction, i.e., we can take $T_1^{\mathrm{J}} = \tau^{\mathrm{L}}$ and $T_2^{\mathrm{J}} = \tau^{\mathrm{R}}$ and furthermore $R^{\mathrm{J}} = |q|/|T_1^{\mathrm{J}} - T_2^{\mathrm{J}}|$. The junction temperatures T_1^{J} and T_2^{J} can be specified at any value by assigning the desired temperatures of the thermal baths. Essentially in this way we can measure the function $R^{\mathrm{J}}(T_1^{\mathrm{J}}, T_2^{\mathrm{J}})$. For the situations with low heat current passing the junction, $R^{\mathrm{J}}(T_1^{\mathrm{J}}, T_2^{\mathrm{J}})$ is reduced to $R^{\mathrm{J}}(T)$ with $T = (T_1^{\mathrm{J}} + T_2^{\mathrm{J}})/2$ ($|T_1^{\mathrm{J}} - T_2^{\mathrm{J}}| \ll T$) as many experiments measured [113]. However, the form of $R^{\mathrm{J}}(T)$ is not applicable to large current cases that we are particularly interested in. We believe that the form of $R^{\mathrm{J}}(T_1^{\mathrm{J}}, T_2^{\mathrm{J}})$ provides us the complete characterizations of the junction at the phenomenological level.

We can then apply $R^{\mathrm{J}}(T_1^{\mathrm{J}}, T_2^{\mathrm{J}})$ measured in this way to the system with a single abrupt junction at x^{J} to calculate the DTC. Similar to the discussions in the previous section, the heat current q is expressed as

$$q = -\kappa_1(x, T_1(x))\frac{\mathrm{d}T_1}{\mathrm{d}x}, \quad x^{\mathrm{L}} < x < x^{\mathrm{J}}; q = -\kappa_2(x, T_2(x))\frac{\mathrm{d}T_2}{\mathrm{d}x}, \quad x^{\mathrm{J}} < x < x^{\mathrm{R}} \tag{2.25}$$

with the last two conditions in Eq. (2.23) and the junction condition

$$T_1^{\mathrm{J}} - T_2^{\mathrm{J}} = qR^{\mathrm{J}}(T_1^{\mathrm{J}}, T_2^{\mathrm{J}}). \tag{2.26}$$

Applying the variation $\delta T^{\mathrm{L(R)}}$ of the boundary temperature at one end, the resulting temperature profiles T_1 and T_2 are varied to $T_1(x) + \delta T_1(x)$ and $T_2(x) + \delta T_2(x)$ respectively. Of course we have

$$\delta T_1(x^{\mathrm{L}}) = \delta T^{\mathrm{L}}, \quad \delta T_2(x^{\mathrm{R}}) = \delta T^{\mathrm{R}}. \tag{2.27}$$

The heat current q is varied to $q+\delta q$. The junction temperatures $T_1^{\rm J}$ and $T_2^{\rm J}$ are varied to $T_1^{\rm J}+\delta T_1^{\rm J}$ and $T_2^{\rm J}+\delta T_2^{\rm J}$ respectively. We then define the following functions

$$U_1^{\rm L(R)} \equiv \frac{\delta T_1}{\delta T^{\rm L(R)}} \ \text{ and } \ U_2^{\rm L(R)} \equiv \frac{\delta T_2}{\delta T^{\rm L(R)}} \tag{2.28}$$

on $[x^{\rm L}, x^{\rm J}]$ and $[x^{\rm J}, x^{\rm R}]$, respectively. They satisfy the following equations

$$\eta^{\rm L(R)} G^{\rm L(R)} = -\kappa_i \frac{{\rm d}U_i^{\rm L(R)}}{{\rm d}x} - \frac{\partial \kappa_i}{\partial T_i}\frac{{\rm d}T_i}{{\rm d}x} U_i^{\rm L(R)}, \ i=1,2 \tag{2.29}$$

and boundary conditions

$$U_1^{\rm L}(x^{\rm L})=1, U_1^{\rm R}(x^{\rm L}) = U_2^{\rm L}(x^{\rm R}) = 0, U_2^{\rm R}(x^{\rm R}) = 1. \tag{2.30}$$

Their solutions are

$$U_i^{\rm L(R)}(x) = \frac{U_i^{\rm L(R)}(x^{\rm L})}{F_i(x)} - \frac{\eta^{\rm L(R)} G^{\rm L(R)}}{F_i(x)} \int_{x^{\rm J}}^{x} \frac{F_i(t)}{\kappa_i(t, T_i(t))} {\rm d}t \tag{2.31}$$

where

$$F_i(x) \equiv \exp\left(\int_{x^{\rm J}}^{x} \frac{1}{\kappa_i(t,T_i(t))} \frac{\partial \kappa_i}{\partial T_i} \frac{{\rm d}T_i}{{\rm d}t} {\rm d}t \right), \ i=1,2. \tag{2.32}$$

The variation of Eq. (2.26) gives

$$U_1^{\rm L(R)}(x^{\rm J}) - U_2^{\rm L(R)}(x^{\rm J}) = \eta^{\rm L(R)} G^{\rm L(R)} R^{\rm J} + q \frac{\partial R^{\rm J}}{\partial T_1^{\rm J}} U_1^{\rm L(R)}(x^{\rm J}) + q \frac{\partial R^{\rm J}}{\partial T_2^{\rm J}} U_2^{\rm L(R)}(x^{\rm J}). \tag{2.33}$$

By evaluating Eq. (2.31) at $x^{\rm L}$ for $i=1$ and at $x^{\rm R}$ for $i=2$, we can calculate $U_1^{\rm L(R)}(x^{\rm J})$ and $U_2^{\rm L(R)}(x^{\rm J})$ and fit them into Eq. (2.33), and finally get the DTC

$$G^{\rm L} = \left(1 - q\frac{\partial R^{\rm J}}{\partial T_1^{\rm J}}\right) \frac{F_1(x^{\rm L})}{R^{\rm D}}, \ G^{\rm R} = \left(1 + q\frac{\partial R^{\rm J}}{\partial T_2^{\rm J}}\right) \frac{F_2(x^{\rm R})}{R^{\rm D}}, \tag{2.34}$$

where

$$R^{\rm D} \equiv R^{\rm J} + \left(1 - q\frac{\partial R^{\rm J}}{\partial T_1^{\rm J}}\right) \int_{x^{\rm L}}^{x^{\rm J}} \frac{F_1(t)}{\kappa_1(t,T_1(t))} {\rm d}t + \left(1 + q\frac{\partial R^{\rm J}}{\partial T_2^{\rm J}}\right) \int_{x^{\rm J}}^{x^{\rm R}} \frac{F_2(t)}{\kappa_2(t,T_2(t))} {\rm d}t. \tag{2.35}$$

If the TCR is independent of the junction temperatures T_1^{J} and T_2^{J}, the terms with the partial derivatives of R^{J} in Eq. (2.34) are zero and Eq. (2.34) is reduced to

$$G^{\mathrm{L}} = \frac{F_1(x^{\mathrm{L}})}{R^{\mathrm{E}}}, G^{\mathrm{R}} = \frac{F_2(x^{\mathrm{R}})}{R^{\mathrm{E}}}, \tag{2.36}$$

with

$$R^{\mathrm{E}} \equiv R^{\mathrm{J}} + \int_{x^{\mathrm{L}}}^{x^{\mathrm{J}}} \frac{F_1(t)}{\kappa_1(t, T_1(t))}\mathrm{d}t + \int_{x^{\mathrm{J}}}^{x^{\mathrm{R}}} \frac{F_2(t)}{\kappa_2(t, T_2(t))}\mathrm{d}t. \tag{2.37}$$

Each term in the numerators and denominators in Eq. (2.36) are positive, thus $G^{\mathrm{L(R)}} > 0$ and there is no NDTC, which tell us that a temperature dependent TCR is necessary for the existence of NDTC.

With the presence of the temperature dependence of TCR, the contour lines defined in the previous section are described by $G^{\mathrm{L}}\mathrm{d}T^{\mathrm{L}} = G^{\mathrm{R}}\mathrm{d}T^{\mathrm{R}}$, i.e.,

$$\left(1 - q\frac{\partial R^{\mathrm{J}}}{\partial T_1^{\mathrm{J}}}\right)\frac{F_1(x^{\mathrm{L}})}{R^{\mathrm{D}}}\mathrm{d}T^{\mathrm{L}} = \left(1 + q\frac{\partial R^{\mathrm{J}}}{\partial T_2^{\mathrm{J}}}\right)\frac{F_2(x^{\mathrm{R}})}{R^{\mathrm{D}}}\mathrm{d}T^{\mathrm{R}}. \tag{2.38}$$

We denote the upper bound of $|R^{-1}\partial R^{\mathrm{J}}/\partial T_{1,2}^{\mathrm{J}}|$ as T_B^{-1}, where R is the thermal resistance of the whole system, including the TCR. Inside the belt defined by $|T^{\mathrm{L}} - T^{\mathrm{R}}| < T_B$ on the quarter plane we have $|q\partial R^{\mathrm{J}}/\partial T_{1,2}^{\mathrm{J}}| < 1$ and subsequently $R^{\mathrm{D}} > 0$ and $G^{\mathrm{L(R)}} > 0$: no NDTC is displayed in this belt of low bias regime. Therefore, high bias outside this belt is another necessary condition for the existence of NDTC.

As the temperature bias ($|T^{\mathrm{L}} - T^{\mathrm{R}}|$) is increasing beyond T_B, we may arrive at the singularity defined by $R^{\mathrm{D}} = 0$ at which we have infinite $G^{\mathrm{L(R)}}$. We will defer our discussion of such singularity and assume $R^{\mathrm{D}} \neq 0$. There are two cases that present NDTC: 1) $1 - q\partial R^{\mathrm{J}}/\partial T_1^{\mathrm{J}}$ and $1 + q\partial R^{\mathrm{J}}/\partial T_2^{\mathrm{J}}$ have different sign and 2) both $1 - q\partial R^{\mathrm{J}}/\partial T_1^{\mathrm{J}}$ and $1 + q\partial R^{\mathrm{J}}/\partial T_2^{\mathrm{J}}$ are negative and R^{D} is positive. If $\partial R^{\mathrm{J}}/\partial T_1^{\mathrm{J}} + \partial R^{\mathrm{J}}/\partial T_2^{\mathrm{J}} \neq 0$ (which holds for most situations), case 1) is the condition for the first onset of NDTC. In this regard, we believe case 1) is the necessary and sufficient condition for the existence of NDTC.

Case 1) is equivalent to the requirement of

$$\left(1 - q\frac{\partial R^{\mathrm{J}}}{\partial T_1^{\mathrm{J}}}\right)\left(1 + q\frac{\partial R^{\mathrm{J}}}{\partial T_2^{\mathrm{J}}}\right) < 0. \tag{2.39}$$

Eq. (2.39) means that the slope (either $\mathrm{d}T^{\mathrm{L}}/\mathrm{d}T^{\mathrm{R}}$ or $\mathrm{d}T^{\mathrm{R}}/\mathrm{d}T^{\mathrm{L}}$) of the contour line is negative. We can write Eq. (2.39) in a more transparent way by rewriting the temperature dependence of the TCR $R^{\mathrm{J}}(T_1^{\mathrm{J}}, T_2^{\mathrm{J}}) = R^{\mathrm{J}}(\hat{T}^{\mathrm{J}}, \bar{T}^{\mathrm{J}})$ where $\hat{T}^{\mathrm{J}} \equiv T_1^{\mathrm{J}} - T_2^{\mathrm{J}}$ and $\bar{T}^{\mathrm{J}} \equiv T_1^{\mathrm{J}} + T_2^{\mathrm{J}}$:

$$\left|q\frac{\partial R^{\mathrm{J}}}{\partial \hat{T}^{\mathrm{J}}} - 1\right| < \left|q\frac{\partial R^{\mathrm{J}}}{\partial \bar{T}^{\mathrm{J}}}\right|. \tag{2.40}$$

Eqs. (2.40) imply that the existence of NDTC requires a certain way of the junction temperature dependence of the TCR. Namely, the TCR must be dependent on $\bar{T}^{\mathrm{J}}$, which is of significant physical relevance, because in the world of thermal transport we have a natural temperature ground of absolute zero temperature. This indicates the drastic difference between the thermal and electrical transport where in the latter case the junction behavior only depends on the voltage drop across the junction.

Case 2) is equivalent to the requirement of

$$\begin{aligned} &q\frac{\partial R^{\mathrm{J}}}{\partial \hat{T}^{\mathrm{J}}} - 1 > \left|q\frac{\partial R^{\mathrm{J}}}{\partial \bar{T}^{\mathrm{J}}}\right|, \\ &|q| < \frac{R^{\mathrm{J}} + \displaystyle\int_{x^{\mathrm{L}}}^{x^{\mathrm{J}}} \frac{F_1(t)}{\kappa_1(t, T_1(t))}\mathrm{d}t + \int_{x^{\mathrm{J}}}^{x^{\mathrm{R}}} \frac{F_2(t)}{\kappa_2(t, T_2(t))}\mathrm{d}t}{\left|\dfrac{\partial R^{\mathrm{J}}}{\partial T_1^{\mathrm{J}}}\right| \displaystyle\int_{x^{\mathrm{L}}}^{x^{\mathrm{J}}} \frac{F_1(t)}{\kappa_1(t, T_1(t))}\mathrm{d}t + \left|\dfrac{\partial R^{\mathrm{J}}}{\partial T_2^{\mathrm{J}}}\right| \displaystyle\int_{x^{\mathrm{J}}}^{x^{\mathrm{R}}} \frac{F_2(t)}{\kappa_2(t, T_2(t))}\mathrm{d}t}. \end{aligned} \tag{2.41}$$

Eq. (2.40) and Eq. (2.41) suggest a lower bound on the heat current for the existence of NDTC such that

$$|q| > \frac{1}{\max\{|\partial R^{\mathrm{J}}/\partial T_1^{\mathrm{J}}|, |\partial R^{\mathrm{J}}/\partial T_2^{\mathrm{J}}|\}} \quad \text{and} \quad |q| > \frac{1}{\min\{|\partial R^{\mathrm{J}}/\partial T_1^{\mathrm{J}}|, |\partial R^{\mathrm{J}}/\partial T_2^{\mathrm{J}}|\}} \tag{2.42}$$

respectively. This again indicates that the existence of NDTC is in the regime beyond low heat current.

From Eq. (2.26) we know that $q = \hat{T}^{\mathrm{J}}/R^{\mathrm{J}}$. We may be tempted to consider Eq. (2.39) as a condition imposed solely on the junction. However, Eq. (2.39) is in essential a condition for the whole system, since the junction temperatures are actually determined by the system parameters. There is one exception that the system is composed of two segments of thermal superconductors, i.e., the materials in $[x^{\mathrm{L}}, x^{\mathrm{J}}]$ and $[x^{\mathrm{J}}, x^{\mathrm{R}}]$ have infinite thermal conductivities (or physically have negligibly less thermal resistance than the TCR). The externally applied temperature drop now is entirely on the junction, and namely we have $T_1^{\mathrm{J}} = T^{\mathrm{L}}$ and $T_2^{\mathrm{J}} = T^{\mathrm{R}}$ and $R^{\mathrm{J}} = R^{\mathrm{J}}(T^{\mathrm{L}}, T^{\mathrm{R}})$. The conditions of the existence of NDTC are $q\partial R^{\mathrm{J}}/\partial T^{\mathrm{L}} > 1$ or $q\partial R^{\mathrm{J}}/\partial T^{\mathrm{R}} < -1$.

At the onset of NDTC, we have either $G^{\mathrm{L}} = 0$ or $G^{\mathrm{R}} = 0$ and the heat current (vs. temperature bias) reaches its first local maximum (for $q > 0$) or minimum (for $q < 0$). Taking the case of $G^{\mathrm{L}} = 0$ as an example, we can calculate the following variation rates at the vicinity of $G^{\mathrm{L}} = 0$:

$$\begin{aligned} &\frac{\delta(T^{\mathrm{L}} - T_1^{\mathrm{J}})}{\delta T^{\mathrm{L}}} = 1 - F_1(x^{\mathrm{L}}),\ \frac{\delta(T_1^{\mathrm{J}} - T_2^{\mathrm{J}})}{\delta T^{\mathrm{L}}} = F_1(x^{\mathrm{L}}) > 0, \\ &\frac{\delta(T_2^{\mathrm{J}} - T^{\mathrm{R}})}{\delta T^{\mathrm{L}}} = G^{\mathrm{L}} \int_{x^{\mathrm{J}}}^{x^{\mathrm{R}}} \frac{F_2(t)}{\kappa_2(t, T_2(t))} \mathrm{d}t. \end{aligned} \tag{2.43}$$

In the NDTC regime of $G^{\mathrm{L}} \lesssim 0$, the temperature increase δT^{L} of T^{L} is distributed over $[x^{\mathrm{L}}, x^{\mathrm{R}}]$ in such a way that the temperature drop over $[x^{\mathrm{J}}, x^{\mathrm{R}}]$ is decreasing with increasing T^{L} (which is the manifestation of NDTC) while the temperature drop over the junction is increasing with T^{L}.

As mentioned before, there possibly exist singularities of the DTCs at $R^{\mathrm{D}} = 0$. The zeros of G^{L} and G^{R} exist respectively at $q_z^{\mathrm{L}} = [\partial R^{\mathrm{J}}/\partial T_1^{\mathrm{J}}]^{-1}$ and $q_z^{\mathrm{R}} = -[\partial R^{\mathrm{J}}/\partial T_2^{\mathrm{J}}]^{-1}$. The zeros may accidentally cancel the singularity, but this cancellation can rarely happen. Even if it happens, this accidence can be easily removed by shifting the

temperature that is supposed to be fixed. If the cancellation does not happen, we have infinite $G^{\mathrm{L(R)}}$ at the singularities.

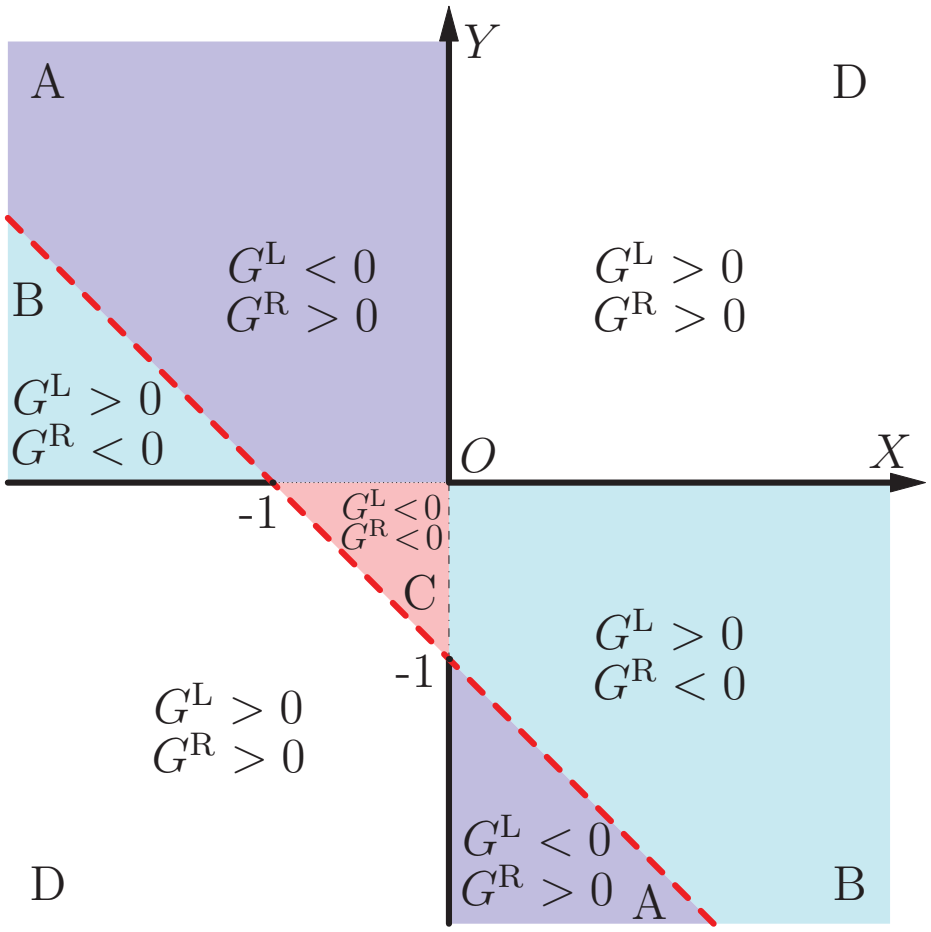

Fig. 2.21. The phase diagram on the $X - Y$ plane that present NDTC.

We denote the dimensionless quantities

$$\begin{aligned} X &\equiv \left(1 - q\frac{\partial R^{\mathrm{J}}}{\partial T_1^{\mathrm{J}}}\right)\frac{J_1}{R^{\mathrm{J}}} = \frac{\partial q}{\partial T_1^{\mathrm{J}}}J_1, \\ Y &\equiv \left(1 + q\frac{\partial R^{\mathrm{J}}}{\partial T_2^{\mathrm{J}}}\right)\frac{J_2}{R^{\mathrm{J}}} = -\frac{\partial q}{\partial T_2^{\mathrm{J}}}J_2 \end{aligned} \tag{2.44}$$

such that

$$G^{\mathrm{L}} = \frac{X}{1+X+Y}\frac{F_1(x^{\mathrm{L}})}{J_1}, G^{\mathrm{R}} = \frac{Y}{1+X+Y}\frac{F_2(x^{\mathrm{R}})}{J_2}, \tag{2.45}$$

to study the phase diagram of the existence of NDTC. Now there exists NDTC if and only if at least one of X and Y is negative, which means that at least one of $\partial q/\partial T_1^{\mathrm{J}}$ and $-\partial q/\partial T_2^{\mathrm{J}}$ is negative [1]. We refer to such junctions as those with *intrinsic junction NDTC* which is now necessary and sufficient for NDTC to occur. Thus, the existence of NDTC in systems with a single abrupt junction is uniquely determined by the properties of the TCRs, regardless of the properties of the system away from the junction.

Furthermore, we can formulate the existence of NDTC on the X-Y plane: find out the points on the plane that correspond to negative G^{L} or G^{R}. NDTC exists inside the shaded areas (A, B and C in Fig. 2.21), not including the boundaries labelled by the thick solid black and red dashed lines. Note that we have $G^{\mathrm{L}} < 0$ and $G^{\mathrm{R}} = 0$ on the thin dotted line ($-1 < X < 0$, $Y = 0$) while $G^{\mathrm{L}} = 0$ and $G^{\mathrm{R}} < 0$ on the thin dash-dotted line ($X = 0$, $-1 < Y < 0$). We have $R^{\mathrm{D}} = 0$ and thus infinite ($\pm\infty$) DTCs on the thick red dashed line. For the points in the shaded areas and close to the thick red dashed line, we can have very large magnitude of NDTC, useful to design sensitive detectors for temperature fluctuations.

The blue and cyan shaded areas A and B in Fig. 2.21 are not bounded on the X-Y plane. They correspond to one of G^{L} and G^{R} is negative and the other is positive, i.e., $G^{\mathrm{L}}G^{\mathrm{R}} < 0$. The red shaded area C in Fig. 2.21 is bounded and it corresponds to

[1] To have NDTC (negative G^{L} or G^{R}), it is necessary that at least one of X and Y is negative. Conversely, if at least one of X and Y is negative, since both X and Y (assumed to be continuous with temperature bias) would be positive in the limit of vanishing temperature bias, there exists a critical temperature bias below which both X and Y are positive and slightly above which at least one of X and Y is negative and $|X| \ll 1$ and $|Y| \ll 1$, thus one of G^{L} and G^{R} must be negative ($1 + X + Y$ is positive) and there exists NDTC.

the case that both $1 - q\partial R^{\mathrm{J}}/\partial T_1^{\mathrm{J}}$ and $1 + q\partial R^{\mathrm{J}}/\partial T_2^{\mathrm{J}}$ are negative and R^{D} is positive. There is no NDTC in the blank area D.

For systems with multiple abrupt junctions, both NDTC and singularities are possible and their detailed occurrence conditions are more complicated. However, these junctions can be grouped into a single junction with its properties determined by the way (e.g., the order and the connection materials) to organize the junctions and by the properties of those individual junctions. This provides us a routine to engineer the TCR. For example, we can construct a system composed of three segments in $[x^{\mathrm{L}}, x_1^{\mathrm{J}}]$, $[x_1^{\mathrm{J}}, x_2^{\mathrm{J}}]$ and $[x_2^{\mathrm{J}}, x^{\mathrm{R}}]$. Suppose that $[x^{\mathrm{L}}, x_1^{\mathrm{J}}]$ and $[x_2^{\mathrm{J}}, x^{\mathrm{R}}]$ contain the same kind of material and the material in $[x_1^{\mathrm{J}}, x_2^{\mathrm{J}}]$ is different. We can effectively have a single junction with its TCR R^{J} characterized by the temperature T_1^{J} at x_1^{J} at the side of $[x^{\mathrm{L}}, x_1^{\mathrm{J}}]$, the temperature T_2^{J} at x_2^{J} at the side of $[x_2^{\mathrm{J}}, x^{\mathrm{R}}]$ and the heat current flowing across the effective junction. In this way, we have $R^{\mathrm{J}}(T_1^{\mathrm{J}}, T_2^{\mathrm{J}}) = R^{\mathrm{J}}(T_2^{\mathrm{J}}, T_1^{\mathrm{J}})$.

2.6 Summary

We have used classical molecular dynamics to study the thermal transport in GNRs. Our predictions include the following aspects.

(1) The calculated thermal conductivity of symmetric rectangular GNRs is on the same order of magnitude as the value expected for graphene but with differences likely caused by the finite sizes of GNRs. The thermal conductivity is shown to depend on edge chirality and the zigzag edge GNRs is shown to have larger thermal conductivity than that of the armchair edge GNRs. We have demonstrated the thermal rectification effect in asymmetric triangular and trapezoidal GNRs. The triangular GNR with vertex angle of 30° and armchair bottom edge is found to have the largest thermal rectification among all GNRs studied. The

defects can reduce the thermal conductivity of GNRs as well as the thermal rectification in asymmetric GNRs.

(2) The edge H-passivation can reduce the thermal conductivity significantly. We show that the thermal conductivity depends on the concentrations of isotopic atoms and their distribution patterns. The isotopic superlattice distributions can reduce the thermal conductivity much more than random distributions.

(3) We have studied the nonlinear thermal transport in rectangular and triangular GNRs under large temperature biases. We find that in short (6 nm) rectangular GNRs, the NDTC exists in a certain range of applied temperature difference. As the length of the rectangular GNR increases, NDTC gradually weakens. In triangular GNRs, NDTC only exists in the thermal current direction from the narrower to the wider end. The ability to tune and control NDTC by temperature parameters and GNR shapes provides potential ways to manage heat and manipulate thermal signals at the nanoscale.

(4) In order to consider the possibility of the existence of NDTC for long GNRs which are believed to work in the diffusive regime, we also study the steady state 1D thermal transport in the diffusive regime without heat sources or sinks. The Fourier's law is applied to derive the formula for the differential thermal conductance. We find that NDTC cannot exist in systems without any abrupt thermal junctions. However, we could have NDTC by introducing temperature dependent thermal contact resistances (TCRs) that are believed to exist. Our predictions provide a novel approach to realize NDTC through careful thermal contact engineering.

3. THERMAL CONDUCTIVITY MEASUREMENT OF GRAPHENE COMPOSITE

Graphene composite materials have drawn much attention recently mainly due to its potential of large scale applications through mass production and abilities to tune their various properties by material engineering [13]. Graphene composites exhibit excellent properties, such as extremely low percolation threshold of electrical conductivity [13]. Similarly, the thermal properties of graphene composites can be tuned by controlling the graphene concentrations. It has been shown that adding graphene flakes to host materials, such as epoxy [114] and nanostructured phase change materials [115] can dramatically increase the thermal conductivity of the composite materials. However, there are few experiments to study the thermal transport in graphene-polystyrene composite which is inherently uniform through chemically compatible processes[1]. It is interesting to investigate the thermal properties of such graphene composites, and in particular, the graphene concentration dependence of thermal conductivities.

3.1 Synthesis of graphene composite

Graphene composites of several different volume concentrations are made from polystyrene and graphene oxide. Graphene oxide is produced using the Hummers method [116]. The uniform mixture of graphene oxide and polystyrene in dimethyl-

[1]Since graphene surface is chemically inert, it needs to be modified prior to the uniform mixture of graphene and host material, which is considered a chemically compatible process.

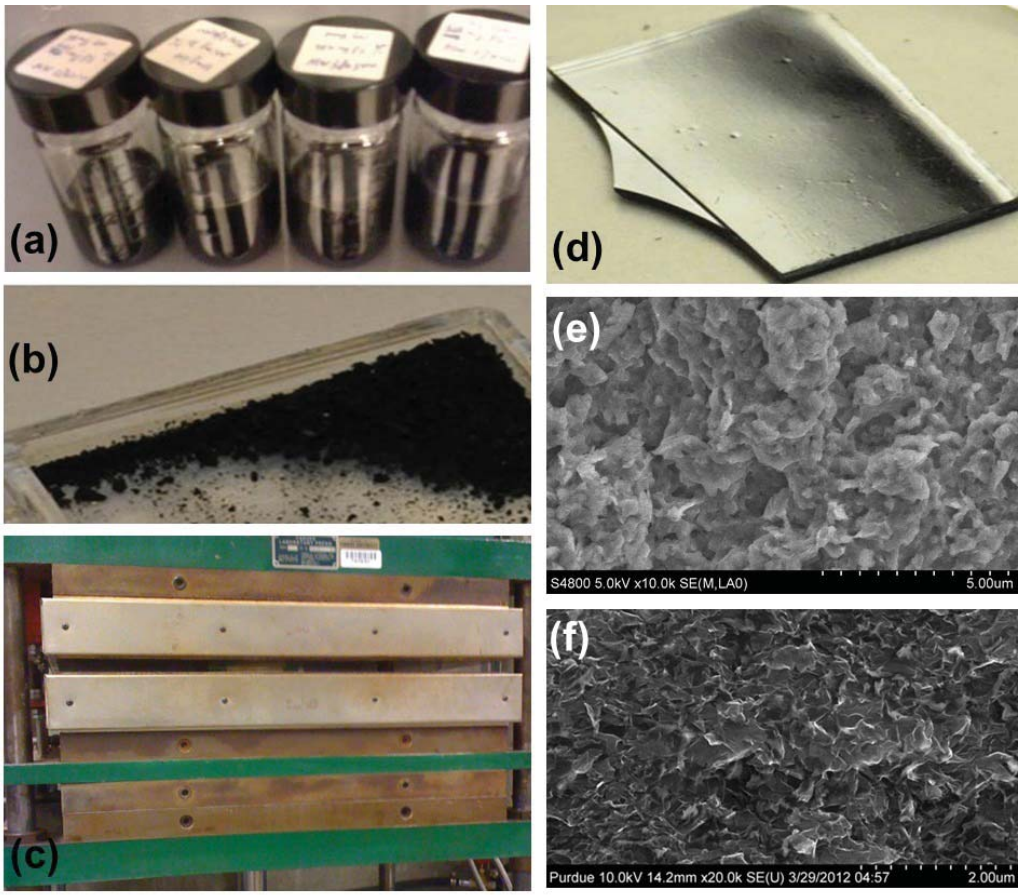

Fig. 3.1. Uniform mixture of reduced graphene oxide and polystyrene in dimethylformamide with different graphene concentrations labeled on the bottles (a), dried graphene composite milled into fine powders (b) and the graphene composite plate (d) produced from the hot press machine (c). The SEM images of graphene oxide and graphene composite (5 vol. %) are shown in (e) and (f) respectively. The thickness of graphene composite plate in (d) is $\sim$ 1.3 mm.

formamide is treated by hydrazine to reduce graphene oxide towards graphene (mostly multilayer graphene flakes) (Fig. 3.1a) [117]. This solution is then polymerized using

methanol and filtered to collect graphene composite, followed by being dried in a vacuum oven. The dried graphene composite is milled into fine powders (Fig. 3.1b) and hot pressed into bulk dense composite plate as shown in Fig. 3.1d. The thickness of the graphene composite plate is about 1.3 mm. Note that the graphene flakes contained in the composite are actually reduced graphene oxide. Without causing confusion, we interchangeably use graphene and reduced graphene oxide.

Our synthesized graphene oxide and composite are characterized firstly by scanning electron microscopy (SEM) to study its uniformity. As shown in Fig. 3.1e, the graphene oxide (note that the graphene oxide produced by Hummer's method is in a solid form) has foam-like structures. From Fig. 3.1f for the graphene composite sample of 5 vol.%, graphene is uniformly dispersed in the polystyrene host. Graphene flakes form the backbone of the umbrella structure in the SEM image [13].

3.2 Electrical characterization of graphene composite

The 4-terminal electrical conductivity of graphene composite is measured[2] using the setup shown in Fig. 3.2a. Shadow masks are used to directly deposit metal contact pads on the surface of graphene composite, followed by manual wire bonding using silver epoxy. During the measurement, a constant DC current ($\sim$ 1 μA) is passing between the two outer contacts, and the voltage drop is measured between the two inner contacts [118]. The electrical conductivity can be calculated from the conductance scaled by the geometrical factors of width w, length l and thickness t.

Fig. 3.2b presents the measured electrical conductivity vs. the graphene volume percent φ. The electrical conductivity exhibit a very strong dependence on φ and

[2]This measurement is led by Wonjun Park in our lab.

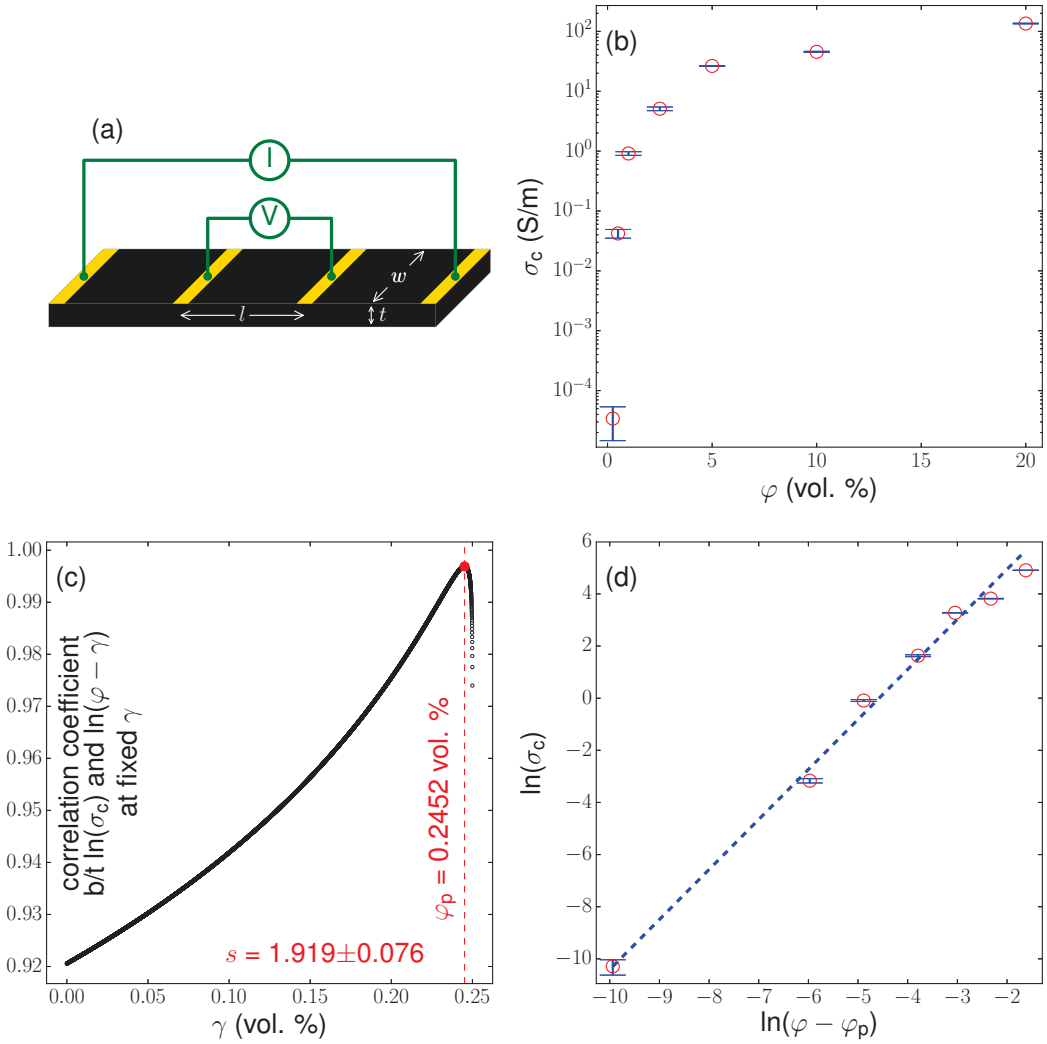

Fig. 3.2. Schematics (a) for measuring the electrical conductivity of graphene composite, electrical conductivity σ_c (b) of graphene composite vs. graphene volume percent φ, correlation coefficient (c) between $\ln(\sigma_c)$ and $\ln(\varphi - \gamma)$ at fixed γ that is less than the minimum volume percent of conducting samples, and the linear fitting (d) of $\ln(\sigma_c)$ vs. $\ln(\varphi - \varphi_p)$ where φ_p is labeled at the red dot in (c).

varies in six order of magnitude. It is reported [13] that such behavior can be fitted to a percolation formula as the following:

$$\sigma_c = \sigma_e(\varphi - \varphi_p)^s \tag{3.1}$$

where σ_e is the effective conductivity of graphene flakes that are embeded inside the composite, φ_p is the percolation threshold (indicating the onset of electrical conduction), and s is the universal critical exponent. Direct fitting for three parameters (σ_e, φ_p and s) using seven data points is challenging and often yields unphysical results (e.g., φ_p can be negative). Thus, an indirect fitting procedure is employed here. All of these three parameters are identified at the maximum point on the curve of the correlation coefficient between $\ln(\sigma_c)$ and $\ln(\varphi - \gamma)$ at fixed γ in Fig. 3.2c. The fitted results are $s = 1.919 \pm 0.076$, $\varphi_p = 0.2452$ vol. % and $\sigma_e = 6494$ S/m. This value of s is from the slope of the linear fit in Fig. 3.2d. This value of s is very close to the universal critical exponent of $s = 2$ for embedding 2D materials in the 3D matrix [13].

3.3 Fundamentals of 3ω method

In general, 3ω method for measuring the thermal conductivity (and thermal diffusivity) refers to extracting the thermal transport properties from the frequency dependent thermal impedance between a heater and the materials to be measured. This thermal impedance is defined as the amplitude of the temperature variation on the heater when unit electrical power generated on (unit length or unit area of) the heater is dissipated through the material. The heater must have nonzero temperature coefficient in order to detect the temperature rise on it by measuring the resistance of the heater. When an AC current at frequency ω is applied on the heater, the Joule heating will induce a temperature rise at frequency 2ω on the heater. Such

temperature rise will induce a variation of the heater resistance at 2ω so that the small voltage at 3ω (corresponding to the 3ω component of the product of the AC current at ω and the resistance component at 2ω) will appear on the heater.

To quantify this procedure, we can assume the AC current is in the form of $i(t) = I_{1\omega}\cos(\omega t)$ at time t. Denote the resistance and the temperature coefficient of the heater as R and β. The temperature rise term of interested on the heater can be assumed to be $\Delta T(t) = T(\omega)\cos(2\omega t + \phi(\omega))$. The appearance of the frequency dependent phase $\phi(\omega)$ is due to the fact that the heat capacity of any material is not zero (except at absolute zero temperature that is not accessible). [physical nature of this phase from the form of the heat equation will be discussed here]

The 3ω component of the voltage $i(t)\beta\Delta T(t)R$ is $v_{3\omega} = \frac{1}{2}\beta R I_{1\omega} T(\omega)\cos(3\omega t + \phi(\omega))$, and thus its amplitude for the in-phase and the out-of-phase components are

$$\begin{aligned} V_{3\omega,\text{in-phase}} &= \frac{1}{2}\beta R I_{1\omega} T_{2\omega,\text{in-phase}} = \frac{1}{2}\beta R I_{1\omega} T(\omega)\cos(\phi(\omega)), \\ V_{3\omega,\text{out-of-phase}} &= \frac{1}{2}\beta R I_{1\omega} T_{2\omega,\text{out-of-phase}} = -\frac{1}{2}\beta R I_{1\omega} T(\omega)\sin(\phi(\omega)), \end{aligned} \tag{3.2}$$

respectively. The average electrical power generated on the heater is $P = \frac{1}{2} I_{1\omega}^2 R$. If we define the complex numbers $V_{3\omega} \equiv V_{3\omega,\text{in-phase}} + jV_{3\omega,\text{out-of-phase}}$ and $T_{2\omega} \equiv T_{2\omega,\text{in-phase}} + jT_{2\omega,\text{out-of-phase}}$, where $j = \sqrt{-1}$ is the unit imaginary number. The thermal impedance is then defined as

$$Z(\omega) \equiv \frac{T_{2\omega}}{P} = \frac{4V_{3\omega}}{I_{1\omega}V_{1\omega}^2\beta} \tag{3.3}$$

where $V_{1\omega} = I_{1\omega}R$ is the amplitude of the voltage on the heater at ω. Since the heater is assumed to be only resistive, the denominator in Eq. (3.3) is a real number while its numerator is a imaginary number. In Eq. (3.3), all the quantities β, $I_{1\omega}$, $V_{1\omega}$ and $V_{3\omega}$ can be conveniently measured, and thus the thermal impedance $Z(\omega)$ is measured.

There are few important length scales: the length and width of the heater, the minimum dimension of the material (e.g., thickness if it is a film) and the thermal

penetration depth $\lambda = \sqrt{\frac{\alpha}{2\omega}}$ where $\alpha = \frac{\kappa}{\rho c_{\mathrm{p}}}$ is the thermal diffusivity. Usually, to simplify the data processing, the samples prepared look like the schematics in Fig. 3.3a where a narrow metal strip is deposited on top of flat surface of a piece of uniform bulk material. If both the length of the heater and the minimum material dimension are much larger than the width of the heater, the calculation of the thermal impedance can be reduced to a 2D problem in the appropriate range of frequency. The thermal penetration depth on one hand must be much smaller than the minimum dimension of the material and the length of the heater; on the other hand, it should be much larger than the thickness of the heater, so that convenient approximations can be taken. This sets the frequency range of the measurement. For example, the thermal diffusivity of polystyrene [119] is around $\alpha = 1.4 \times 10^{-7}$ m^2/s and take the minimum dimension and the thickness of the heater as 1 mm and 150 nm respectively. Setting λ within the range of 750 nm and 0.2 mm leads to the frequency range of 0.3 Hz to 20000 Hz. Our measurements on graphene composites are within the range of 1 Hz to 1000 Hz.

The approximations we take are the following. First of all, the thickness of the heater is neglected, since it is usually about 100 nm, smaller than any dimensions above, in the frequency range of interested. Secondly, the thermal boundary resistance between the heater and the material is neglected, so that the temperature on the heater is taken as the same as the temperature of the material right beneath the heater. This assumption can be verified as the following. The thermal boundary resistance between the heater and the material (glass or graphene composite) will not exceed [120] Thirdly,

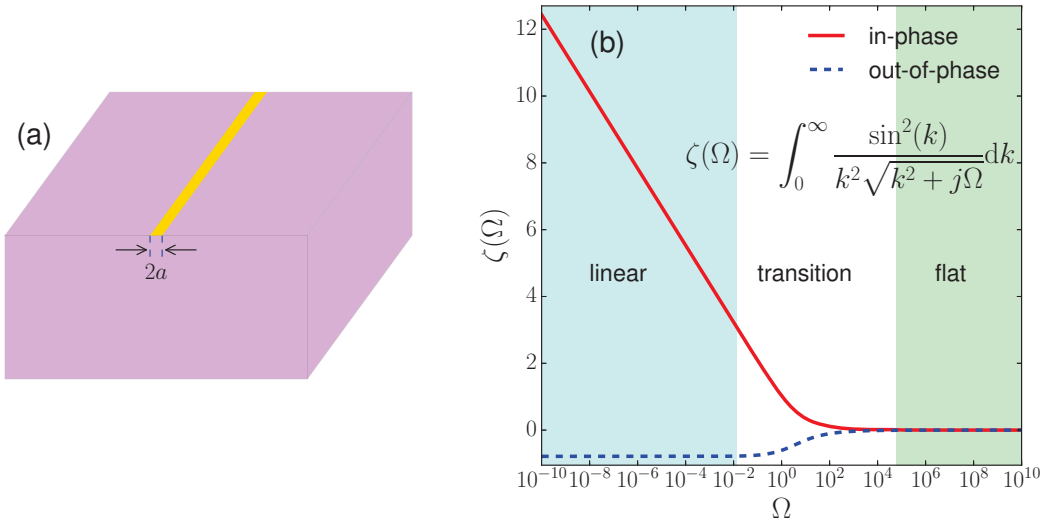

Fig. 3.3. Schematics (a) of a sample of bulk material with a thin and narrow (width of $2a$) metal line on surface and the universal curve (b) of thermal impedance for (a).

The thermal impedance for the sample configuration shown in Fig. 3.3a is $Z(\omega) = \frac{1}{\pi L\kappa}\zeta(\Omega)$ [29, 121] where

$$\zeta(\Omega) = \int_0^\infty \frac{\sin^2(k)}{k^2\sqrt{k^2+j\Omega}}\mathrm{d}k, \tag{3.4}$$

L is the length of the heater, κ is the thermal conductivity of the material and $\Omega = \frac{2\omega a^2}{\alpha} = a^2/\lambda^2$ is the reduced frequency. The curve $\zeta(\Omega)$ is plotted in Fig. 3.3b and three regions of linear (cyan), transition (blank) and flat (green) are identified. The in-phase component of $\zeta(\Omega)$ is linearly (not) dependent on $\ln(\omega)$ in the linear (flat) region, while the out-of-phase component is not dependent on $\ln(\omega)$ in both the linear and flat regions. In the transition region, both the in-phase and out-of-phase varies appreciably with $\ln(\omega)$. As we will see later, our measured data on graphene

composite fall in the transition region where the fitting to the full form of Eq. (3.4) is desired.

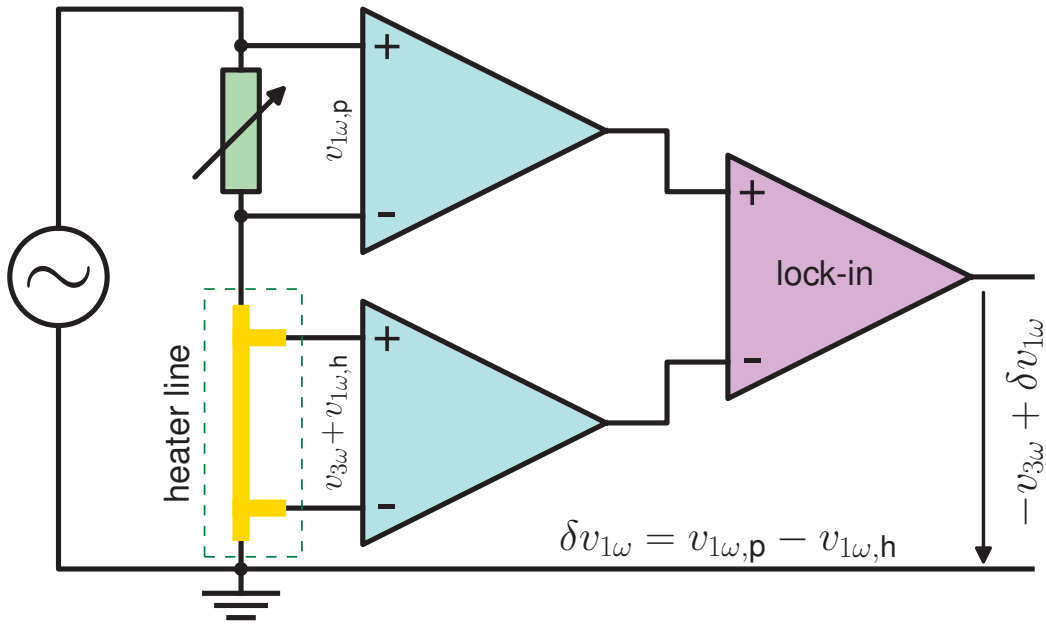

Fig. 3.4. Electrical bridge used to measure the 3ω voltage. The yellow metal line with four terminals enclosed in the green dashed box represents the heater line on the sample surface.

Note that the appearance of the extensive quantity L in $Z(\omega) = \frac{1}{\pi L\kappa}\zeta(\Omega)$ is due to the appearance of the extensive quantity of power P in Eq. (3.3) that is linearly dependent on the length L of the heater when the amplitude $I_{1\omega}$ of the AC heating current is kept a constant. Actually from Eq. (3.3) we can immediately tell $Z(\omega) \propto L^{-1}$ since both $V_{1\omega}$ and $V_{3\omega}$ are proportional to L.

During the measurement, first, the temperature coefficient of the heater is measured for calibration in later data fitting. The sample is then placed in a vacuum stage with pressure typically less than 10^{-5} mTorr. An electrical bridge shown in

Fig. 3.4 is used to cancel the signal $v_{1\omega,\mathrm{h}} = i(t)R$ on the heater at frequency ω by adjusting the potentiometer so that $\delta v_{1\omega} = v_{1\omega,\mathrm{p}} - v_{1\omega,\mathrm{h}}$ appearing at the input of lock-in is as small as possible, where $v_{1\omega,\mathrm{p}}$ is the voltage drop on the potentiometer at frequency ω. Note that the output from the lock-in at the third harmonics is the negative of the 3ω signal $v_{3\omega}$.

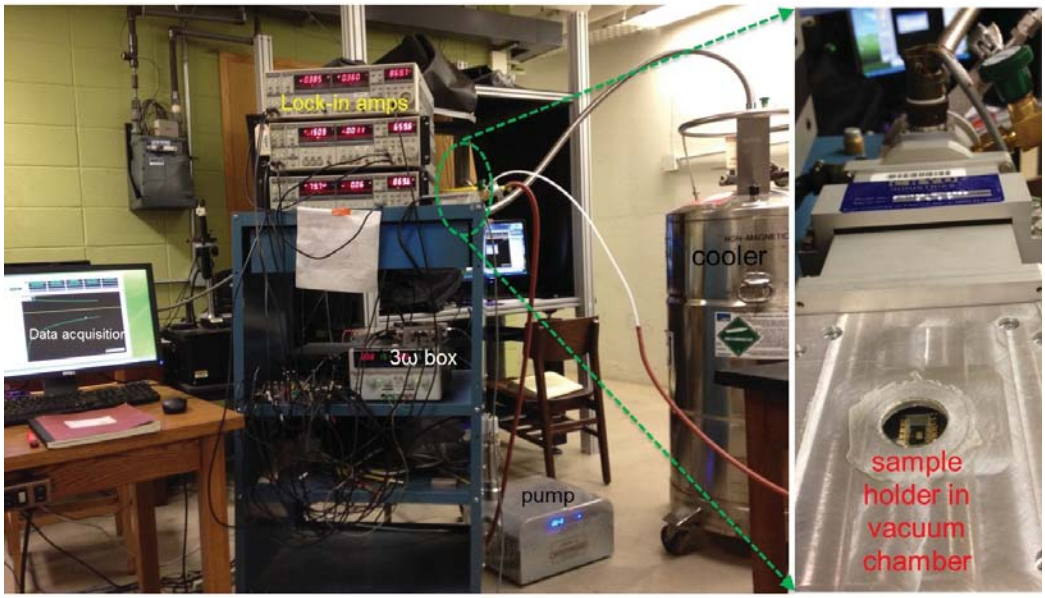

Fig. 3.5. 3ω measurement facilities. The vacuum chamber in which the samples are measured is shown inside the dashed green ellipse and its top view is on the right side. The other parts labeled are lock-in amplifiers, 3ω circuit (Fig. 3.4) box, helium dewar and computer program for data acquisition.

Our measurement set up is shown in Fig. 3.5. It includes the measurement box containing the circuit in Fig. 3.4, lock-in amplifiers used for sensing and detecting the AC signal, cryogenic liquid helium, data acquisition systems, and a cryostage holding

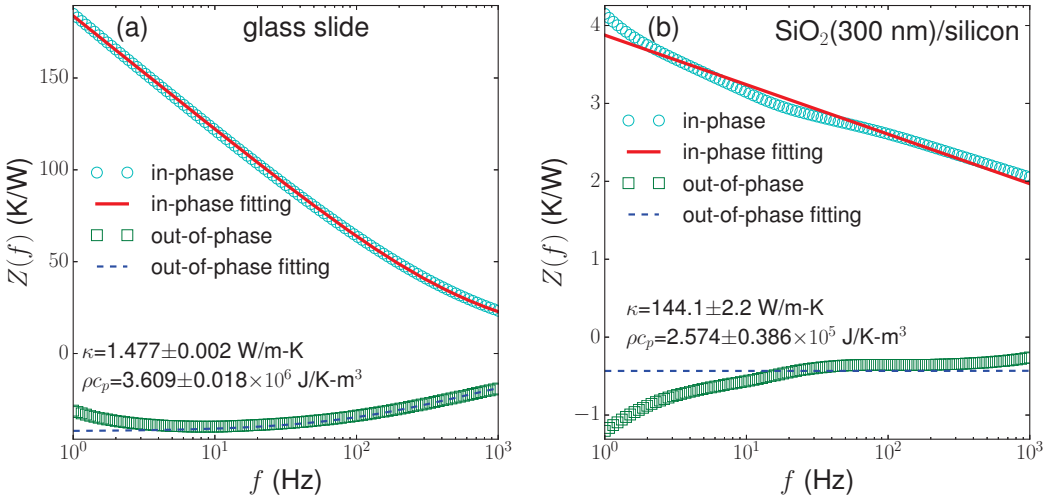

Fig. 3.6. Benchmark of 3ω measurement on glass slide (GOLD SEAL®) and silicon.

the samples in the vacuum chamber. The cryostage is on the right side of by zooming in the dashed green ellipse in the left image.

We benchmark our measurement setup and procedures on a laboratory glass slide (1 mm thick, GOLD SEAL®) and a silicon wafer (0.5 mm thick, with 300 nm SiO_2). The measured thermal impedance is shown in Fig. 3.6. We fit the in-phase data to the in-phase component of $Z(\omega) = \frac{1}{\pi L \kappa}\zeta(\Omega)$ to obtain κ and α and then bring these two parameters back to $Z(\omega) = \frac{1}{\pi L \kappa}\zeta(\Omega)$ to calculate the out-of-phase component. As we can see from Fig. 3.6, the calculated and measured out-of-phase components agree excellently for frequency above few Hz. The measured thermal conductivities for glass slide and silicon are 1.477±0.002 W/m-K and 144.1±2.2 W/m-K respectively. They agree well with the reported values [122, 123] and thus validate our measurement systems.

3.4 3ω characterization of graphene composite

It is challenging to fabricate metal line heaters on the surface of polystyrene based graphene composite materials using the standard lithography techniques, since polystyrene cannot survive in photo-resist strippers such as acetone. In order to avoid the detrimental effect to graphene composite during lithography processes such as lift-off for fabricating metal heaters, a novel transfer method is developed to place the metal heater on the surface of graphene composite. The transfer processes are shown in Fig. 3.7. We start from a piece of silicon substrate covered with polyvinyl alcohol (PVA) film of $\sim$0.5 μm thick (Fig. 3.7a). The standard electron beam lithography is used to fabricate metal lines with contact pads on top of the PVA layer (Fig. 3.7b) and subsequently the substrate is coated with a thin layer of poly(methyl methacrylate) (PMMA) of $\sim$0.3 μm thick. The substrate is then placed in a water bath to remove the PVA layer (Fig. 3.7d) so that the PMMA layer with metal lines is floating on the surface of the water bath as shown in the red dashed circle in Fig. 3.7g. The PMMA layer is then transferred onto the surface of graphene composite (Fig. 3.7e and (3.7 h)). The sample is then soaked in acetone for five seconds to remove the PMMA layer (Fig. 3.7f). It is worthy to note that soaking graphene composite in acetone for five seconds will not damage it, but most of the PMMA layer is stripped off to facilitate the wire bonding. The metal lines are finally bonded onto a chip carrier to perform measurement (Fig. 3.7i).

In Fig. 3.8a-e, the thermal impedance for six graphene composite samples with different graphene volume percent (labeled after the figure label) and the fitted value of κ and ρc_{p} are shown.The measured thermal impedance fits to the formula $Z(\omega) = \frac{1}{\pi L \kappa}\zeta(\Omega)$ well for all the samples. The fitted thermal conductivity κ and the product of mass density and specific heat capacity ρc_{p} vs. graphene volume percent are

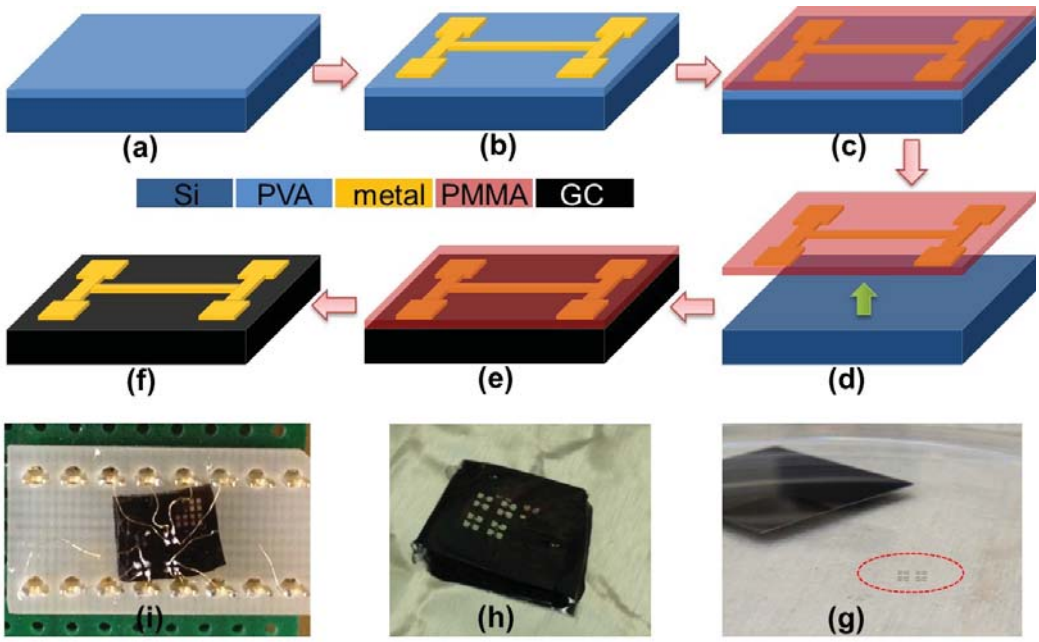

Fig. 3.7. Schematics (a-f) of the transfer procedure to place micrometer sized heater on top of graphene composites and the optical images (g-i) of some steps for the transfer.

shown in Fig. 3.8f which is similar and consistent with our comprehensive study [118]. The enhancement of the thermal conductivity is almost linearly dependent on φ, while ρc_{p} vs. φ does not exhibit simple behavior. The thermal conductivity is increased by $\sim 35\%$ after increasing graphene volume percent to 5 vol.%. Compared to the recent experiment with a much larger enhancement of thermal conductivity of 5 vol.% mechanically exfoliated graphene and epoxy composite [114], we have much less enhancement. This is due to the poor quality of chemically reduced graphene with typical thermal conductivity less than 10 W/m-K [124, 125] which is much less than that of high quality exfoliated graphene.

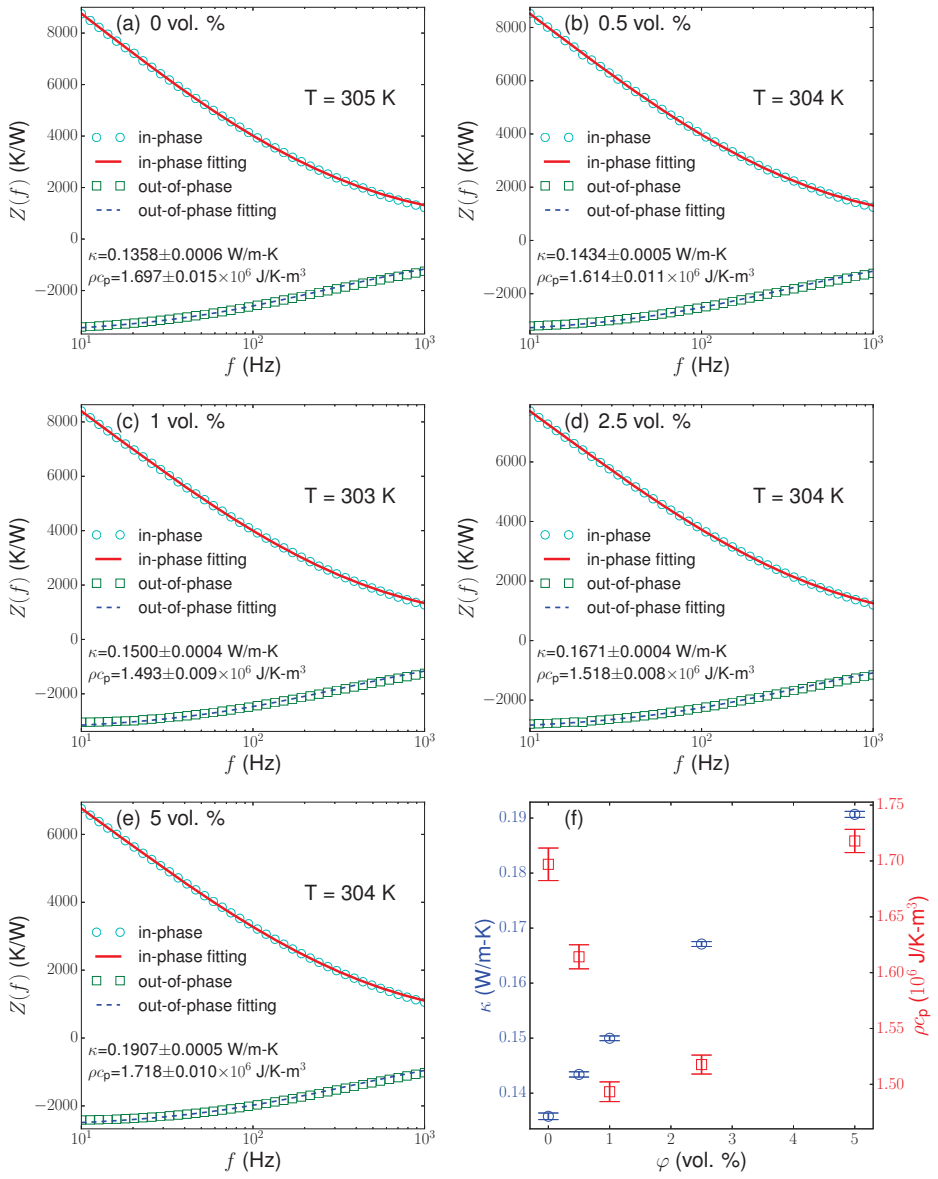

Fig. 3.8. Measured thermal impedance (circles and squares) and fittings (solid lines) for six samples of different graphene volume percent φ (a-e) and the fitted κ and ρc_{p} vs. φ (f).

The pure polystyrene we measured has a thermal conductivity of 0.136 W/m-K which is very close to the reported value [119,126]. The fitted product of mass density and specific heat capacity of pure polystyrene is $C = 1.70 \times 10^6$ J/K-m^3. Given the mass density of 1.05 g/cm^3 [13] of polystyrene used in our experiment, its specific heat capacity is 1.62 J/K-g, close to the reported value of ~ 1.2 J/K-g [127].

The same measurement procedure are also applied to samples of 10 vol. % and 20 vol. %. However the fitted results are an order different from that in Fig. 3.8f. It is found that the temperature coefficient or the (two)four-terminal resistance of the heater is quite different from that on samples with $\varphi \leq 5$ vol. %, possibly due to the electrical leak through the conductive graphene composite. This could happen either at the contact pads or on the heater strip. The results are therefore not presented. This issue is later resolved by Wonjun Park who is taking over the project, and a more systematic study can be found in Ref. [118].

3.5 Summary

We have synthesised graphene composite using chemically reduced graphene as filling and polystyrene as matrix materials. The dried composite is then hot pressed into bulk plate for measurement. We have prepared samples with different graphene filling concentrations. We have developed a novel transfer method to place micrometer sized heater/sensor on composite surface to avoid damaging the composite materials during traditional lithography processes. The measured thermal conductivity is increasing with the filling fraction of graphene. The enhancement can be as high as 50% for 5 vol.% filling.

4. TRANSPORT MEASUREMENT OF GRAPHENE DOUBLE LAYER

Graphene and boron nitride (BN) based heterostructures have attracted intense attention in recent years as the quality of graphene is greatly improved when BN is used as an insulating high quality dielectric material down to 1 nm thick [30, 31, 38, 128, 129]. With the flake transfer technique [30, 36–38] that enables convenient stacking of two dimensional layered materials of a large variety, high quality samples of graphene/BN/graphene have been fabricated to study the interlayer interactions such as Coulomb drag [130] and other novel effects [131–134]. We have built a flake transfer stage (see the stage system and flake transfer recipes in Appendix B) and have fabricated samples of stacked BN/graphene/BN/graphene/BN structures to study the Coulomb drag and counterflow thermoelectric transport.

The electron-hole symmetry allows each layer of graphene to be populated with electrons or holes using metal gates, which is usually difficult to achieve in the traditional semiconductor quantum wells [135]. In the regime where one graphene layer is p-type while the other is n-type, the intriguing exciton condensation [39] is predicted to occur, though the predicted transition temperature spans a wide range [40–49]. This novel phase transition can be detected by performing counterflow transport (equal magnitude and opposite direction of current in two layers), Coulomb drag and other measurements [39, 50]. In addition, thermoelectric transport is a powerful tool for studying the single particle transport [136], strongly correlated systems [137, 138] and macroscopic quantum coherence [139–141] like exciton condensation. Neverthe-

less, no Coulomb drag in bilayer-graphene (BLG) double layers and no counterflow thermoelectric transport measurement have been performed.

4.1 Field effect in bilayer graphene double layer

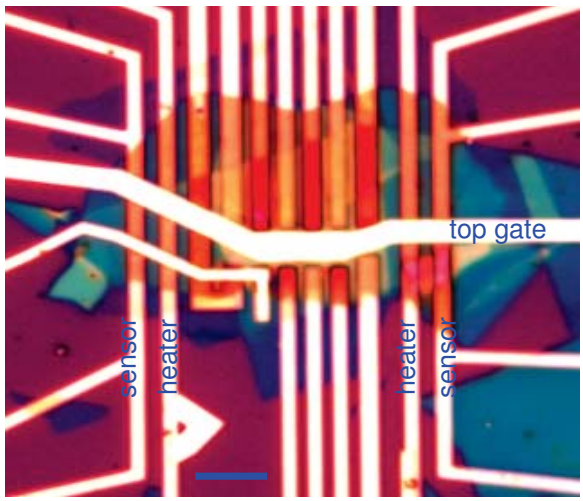

Fig. 4.1. Optical image of the bilayer graphene double layer device. The metal lines for the top gate, heaters, temperature sensors are labeled. The other metal lines are the electrical connection to the Hall bars of both graphene layers. The scale bar is 5 μm.

In this section, we present the data measured from one of our bilayer graphene double layer devices as shown in Fig. 4.1. The existence of the top and bottom gates conveniently allows us to measure the intralayer resistivity for both layers simultaneously, without worrying about the leak between two graphene layers. We can actually simply connect two layers together or disconnect them, both of which will yield the same result. The electrical connection for the device in Fig. 4.1 for the gate voltages $V_{\rm tg}$ and $V_{\rm bg}$ is shown in Fig. 4.2. The carrier type for the top (bottom) graphene layer

can be p(n)-type if $V_{\rm tg}$ ($V_{\rm bg}$) is negative (positive), also indicated by the red filling of the parabolic band structures and the symbols $\oplus$ for holes and $\ominus$ for electrons.

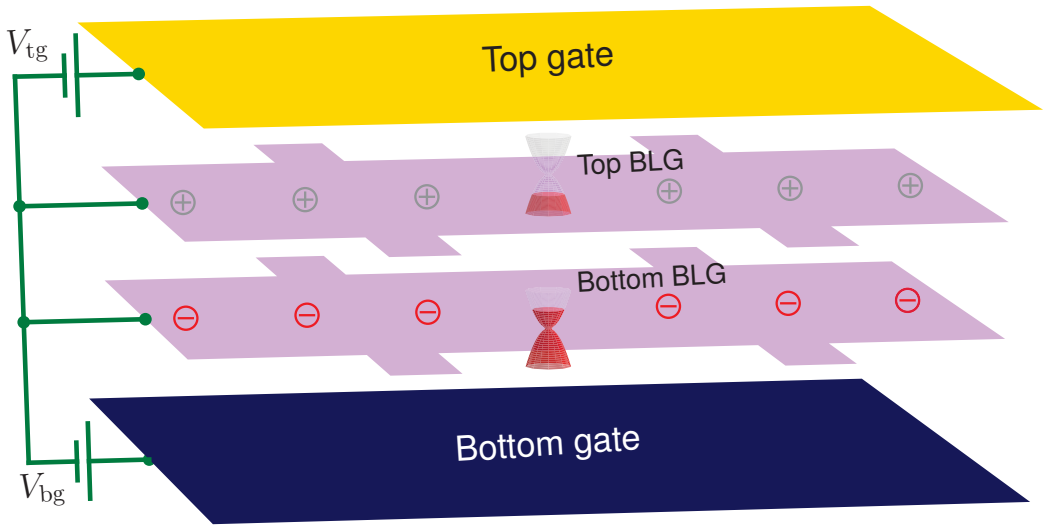

Fig. 4.2. Setup of electrical connection for field effect measurement in graphene double layer.

We have also measured few double layer devices without top gate where we need to apply mutual gate as well as bottom gate voltages. In this case the mutual gate is often leaky if the measurement instruments (e.g., lock-in amplifiers) are directly connected to both layers. Transformers need to be used for at least one graphene layer for electrical isolation (see Appendix E for details of the measurement setup).

Except the Raman spectroscopy, the bilayer nature of both graphene layers in Fig. 4.1 is most convincingly demonstrated by the measured quantum Hall plateaus in high magnetic field, as shown in Fig. 4.3. The longitudinal ($R_{\rm xx}$) and transverse ($R_{\rm xy}$) resistance vs. top ($V_{\rm tg}$) or bottom ($V_{\rm bg}$) and mutual ($V_{\rm mg}$) gate voltages show dips and plateaus respectively and the corresponding filling fractions ν are labeled.

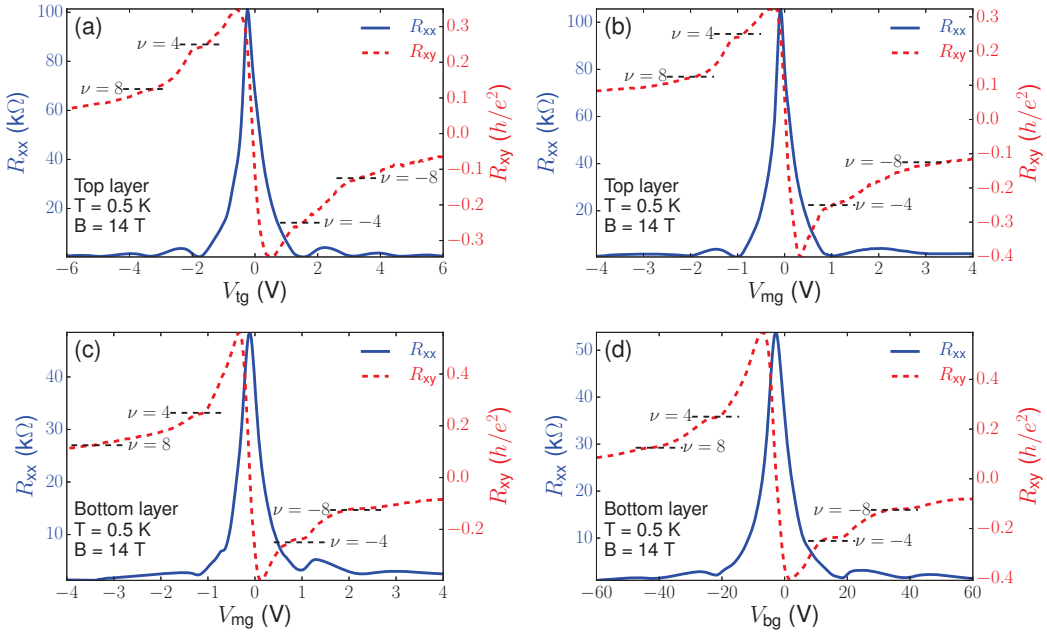

Fig. 4.3. High field magnetotransport of individual BLG layers vs. top (a), bottom (d) and mutual (b-c) gate voltages. The quantum Hall plateaus (dashed horizontal line segments) and the corresponding filling fractions (ν) are labeled.

These labels indicate that the observed quantum Hall plateaus are consistent with theory and experiments. The measured mobility for the bottom (top) graphene layer is around ~ 10000 cm^2/V-s (~ 7000 cm^2/V-s), as shown in Appendix D.1.

The thickness of the boron nitride layer between the two graphene layers is $\sim$ 20 nm. The resulting interlayer tunneling resistance is larger than 10 GΩ, an indication of good electrical isolation between the two graphene layers. The field effect measurement at room temperature and zero field is shown in Fig. 4.4. The top layer resistivity ρ_t vs. V_{tg} and the bottom layer resisitivity ρ_b vs. V_{bg} are shown in (a) and (b) respectively. The colormaps of ρ_t and ρ_b vs. V_{tg} and V_{bg} are shown in (c) and (d)

respectively. The charge neutral points for both graphene layers are very close to zero gate voltage, since both graphene layers are sandwiched between two boron nitride layers, resulting in reduced charge doping of the graphene layers. It is obvious that the resistivity of the top (bottom) graphene layer cannot be effectively tuned by the bottom (top) gate voltage due to the strong screening by the bottom (top) graphene layer. The charge neutral point V_{D} and the maximum resistivity ρ_{D} at V_{D} calculated from (c) and (d) are shown in (e) for the top layer graphene and in (f) for the bottom graphene respectively. The charge neutral points for both graphene layers have a slight negative shift due to the incomplete screening from the other graphene layer. For example, for negative V_{bg}, the incomplete screening of the bottom graphene layer will induce additional holes which need to be compensated by a positive increment of V_{tg}, i.e., $V_{\mathrm{D,top}}$ decreases with increasing V_{bg}. The maximum resistivity vs. gate voltage shows opposite trend for the top and bottom graphene layer. No convincing explanation can be provided for this difference.

The colormaps of the intralayer resistivities and the calculated charge neutral point and maximum resistivity are shown in Fig. 4.5 and Fig. 4.6 respectively for temperatures of 200 K, 100 K and 50 K. The behavior of temperature dependence is more clearly shown in Fig. 4.6. The qualitative behavior of the charge neutral point vs. gate voltage is same for both graphene layers at all temperatures we measured. As temperature drops, the maximum resistivity and the sharpness of the curve of ρ_{b} vs. V_{tg} for the bottom graphene layer exhibits increasingly significant dependence on V_{tg}, while they are weakly dependent on temperature for the top layer graphene.

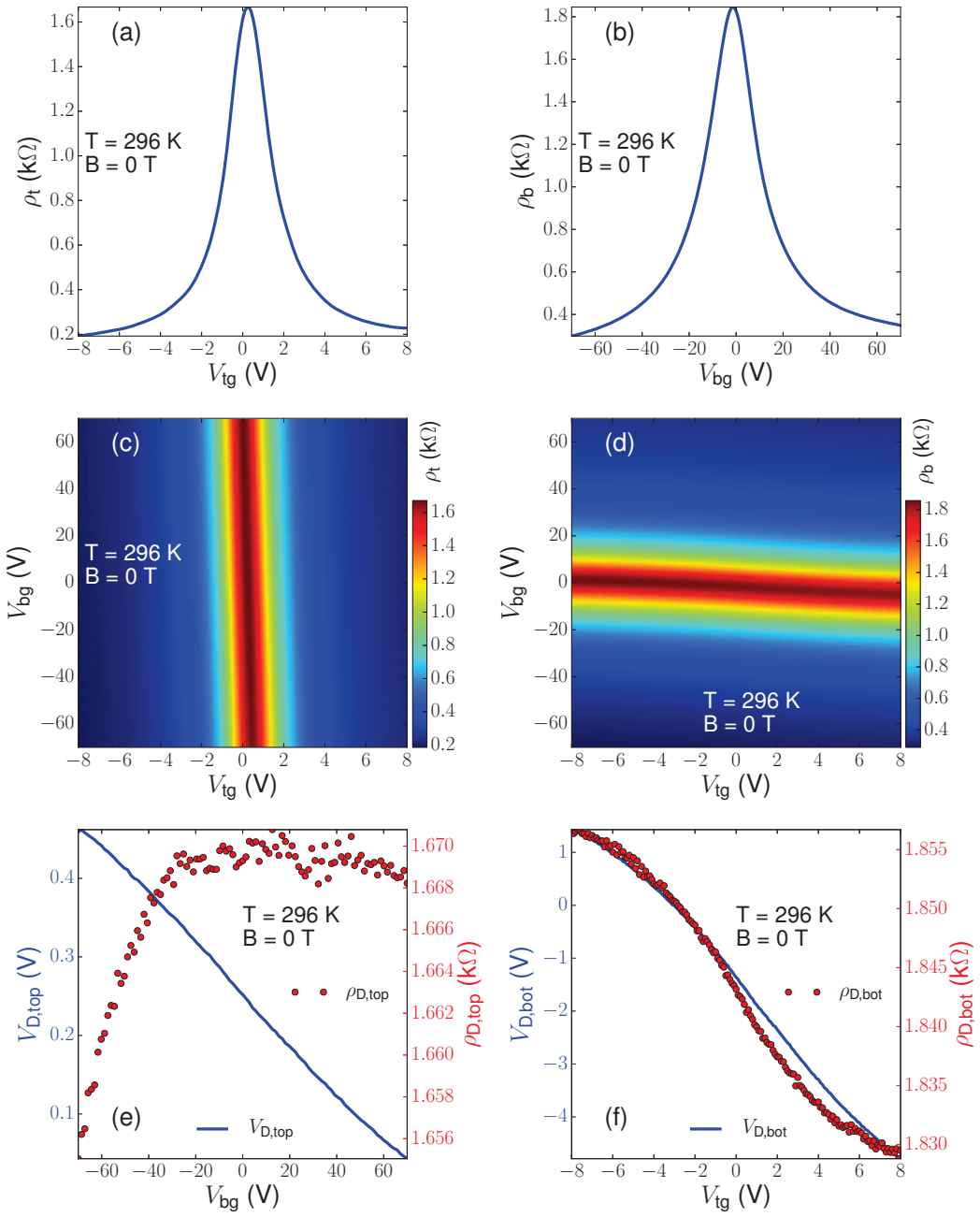

Fig. 4.4. Room temperature field effect of top (a, c, e) and bottom (b, d, f) BLG layer. The charge neutral point (subscript D) V_D and the maximum resistivity ρ_D (e-f) are calculated from the colormaps (c-d) of the resistivity.

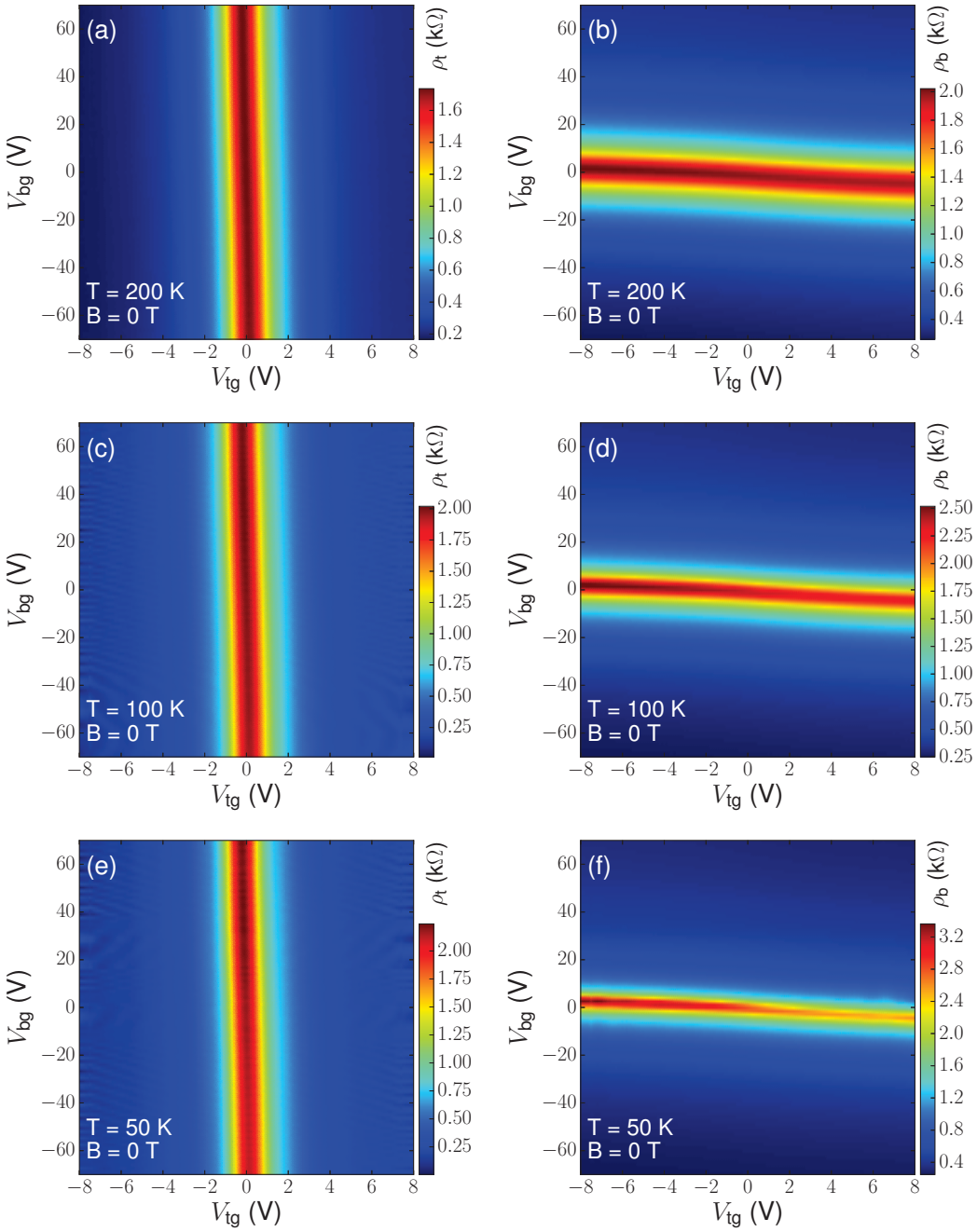

Fig. 4.5. Colormaps of top (a, c, e) and bottom (b, d, f) BLG layers at temperatures of 200 K (a-b), 100 K (c-d) and 50 K (e-f).

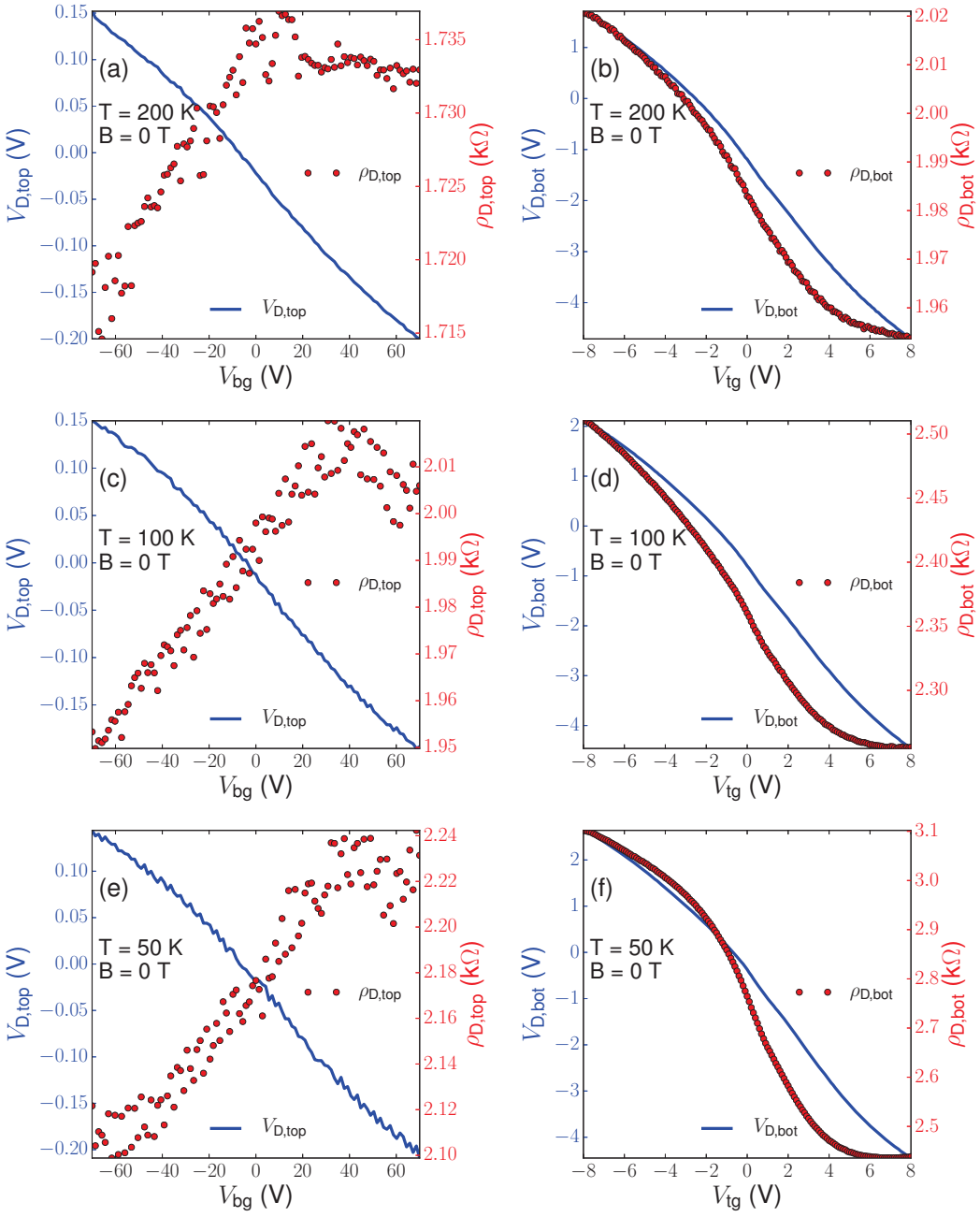

Fig. 4.6. The charge neutral point (subscript D) V_D and the maximum resistivity ρ_D of top (a, c, e) and bottom (b, d, f) BLG layers at temperatures of 200 K (a-b), 100 K (c-d) and 50 K (e-f) calculated from the colormaps in Fig. 4.5.

4.2 Coulomb drag of graphene double layer

Measurement of Coulomb drag resistivity is usually nontrivial and thus special attention should be paid. For example, the magnitude of Coulomb drag resistivity is often quite small, thus lock-in amplifiers are required to measure it. On the other hand, the Coulomb drag resistivity is non-local, which can generate fictitious result when AC signals are applied and detected. We carry out a few precaution steps presented in Appendix D.2 in order to ensure the validity of our measurement.

The Coulomb drag resistivity is defined to be positive when the voltage in the drive layer has the same sign as that in the drag layer when the drag layer has zero current. The colormap of Coulomb drag resistivity ρ_{d} vs. V_{tg} and V_{bg} is shown in Fig. 4.7a. The in-phase and out-of-phase components of ρ_{d} are drawn within the same vertical axis in Fig. 4.7b, corresponding to the horizontal cut at $V_{\mathrm{bg}} = -10$ V in Fig. 4.7a. The out-of-phase component is indeed negligible. The labels (x,y) in Fig. 4.7a-b indicate that the charge carriers of the top layer graphene is of x-type while that of the bottom layer graphene is of y-type where x and y can be either n or p. Though the shape of the colormap is not symmetrical (probably due to charge inhomogeneity), the sign of ρ_{d} is correct, i.e, positive for opposite type of charge carriers in both graphene layers (x$\neq$y) and negative for same type of charge carriers (x=y). Note that the regions of positive ρ_{d} located at the top left and bottom right quadrants are connected across the charge neutral point in both graphene layers, similar to the measurement of Coulomb drag between two single layer graphene (SLG) layers [130]. For another sample of bilayer graphene double layers, the colormap of Coulomb drag resistivity in Fig. 4.8 exhibits different behavior around the charge neutral point in both bilayer graphene layers but similar to the reported result for SLG double layers [142].

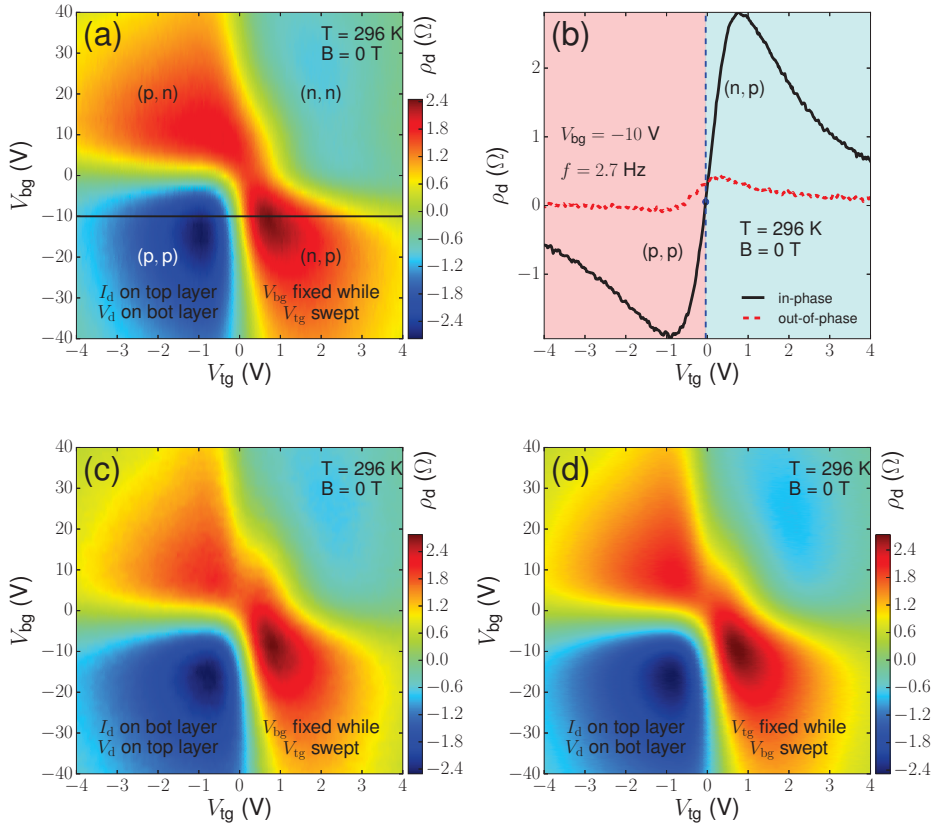

Fig. 4.7. Colormap (a) of Coulomb drag resistivity ρ_d vs. V_{tg} and V_{bg}, (b) ρ_d vs. V_{tg} when $V_{bg} = -10$ V. Note that (b) is corresponding to the horizontal cut in (a), and colormaps of ρ_d measured at the same condition as (a) except that the drive and drag layers are swapped (c) and the gate voltage being swept is V_{bg}. The vertical blue dashed line in (b) divides it into red and cyan regions of p- and n-type of top layer graphene. The labels (x,y) in (a) and (b) indicate the x(y)-type of charge carriers for top (bottom) layer graphene.

The colormaps of Coulomb drag resistivity ρ_d at T = 240 K and T = 200 K are shown in Fig. 4.9a and Fig. 4.9b respectively. The data below 200 K is not collected due to small signal to noise ratio of the Coulomb drag signal. The measured data

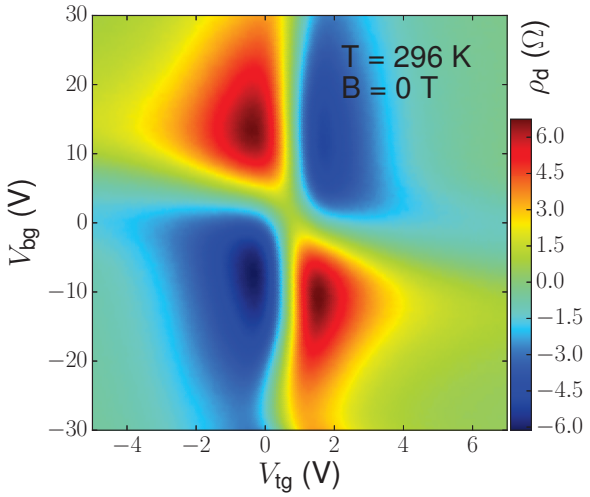

Fig. 4.8. Colormap of Coulomb drag resistivity ρ_d vs. V_{tg} and V_{bg} for another sample of bilayer-graphene double layers measured at room temperature.

exhibits the anomalous behavior that the magnitude of the red peak of ρ_d increases when the temperature is reduced. This is possibly due to the sample inhomogeneity. A representative curve of $|\rho_d|$ vs. T at $V_{tg} = -3.4$V and $V_{bg} = -20$V is shown in Fig. 4.9c, and its log-log plot is shown in Fig. 4.9d where the exponent of α in the relation $|\rho_d| \sim T^\alpha$ is labeled.

The correlation coefficient (c.c.)[1] between $\log(|\rho_d|)$ and $\log(T)$ is calculated for all the gate voltages, and its colormap is shown in Fig. 4.9e. The magnitude of c.c.

[1]The c.c. between two sets of data, e.g., $\{x_1, x_2, \cdots, x_n\}$ and $\{y_1, y_2, \cdots, y_n\}$ is defined as

$$\text{c.c.} = \frac{\sum_{i=1}^{n} |(x_i - \bar{x})(y_i - \bar{y})|}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

where

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}, \ \bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}.$$

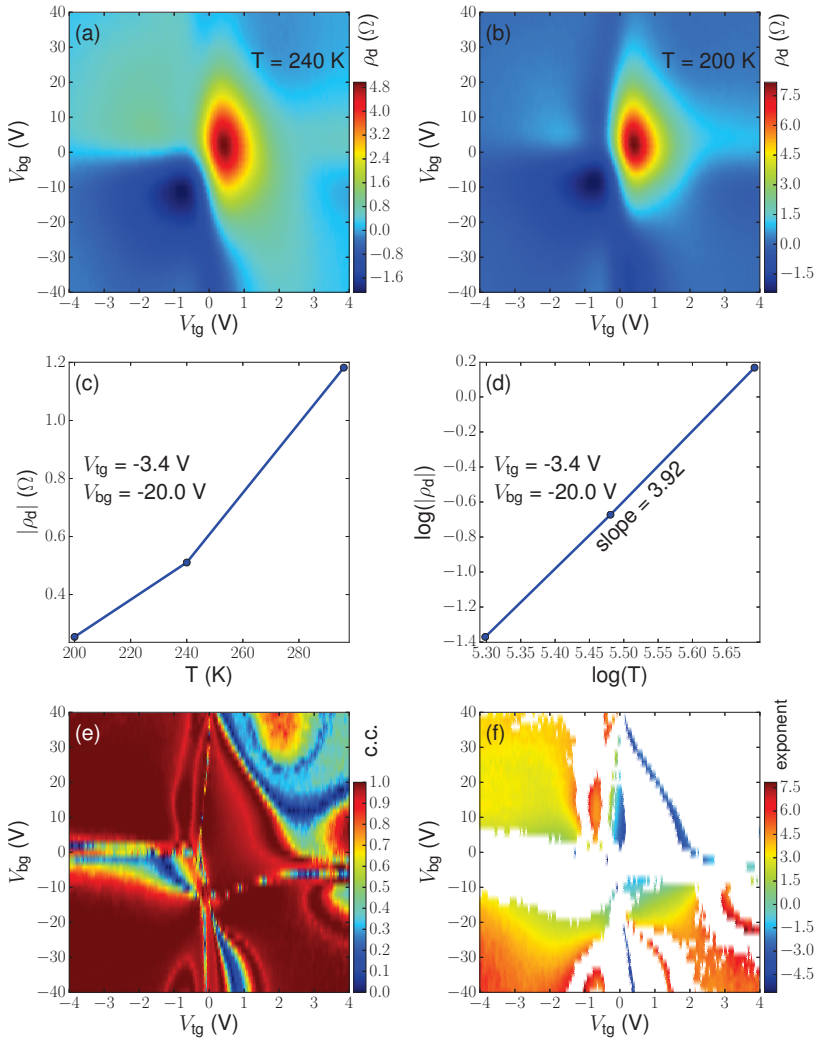

Fig. 4.9. (a) ρ_d at T = 240 K, (b) ρ_d at T = 200 K, (c) $|\rho_d|$ vs. T for V_{tg} = -3.4 V and V_{bg} = -20.0 V, (d) the log-log plot of (c) and the slope indicate the exponent α in the relation $|\rho_d| \sim T^{\alpha}$, (e) correlation coefficient (c.c.) between log($|\rho_d|$) and log(T) and (f) the exponent α in the relation $|\rho_d| \sim T^{\alpha}$. The blank area in (f) corresponds to the gate voltages at which c.c. in (e) is less than 0.99.

measures the linearity between the two data sets: c.c. = 1 implies that the two sets are perfectly linearly dependent. The colormap of the exponent α in the relation $|\rho_{\rm d}| \sim {\rm T}^{\alpha}$ is shown in Fig. 4.9f where the blank area corresponds to the gate voltages at which c.c. is less than 0.99. The exponent α spans in a range of -5 to 7.5. Generally in the Fermi liquid regime, the exponent α is expected to be 1 or 2 in some limiting cases [143]. The behavior of $\rho_{\rm d}$ vs. T could be complicated in the other regimes. For the samples we measured, the parameters in Ref. [143] are $\bar{d} \sim 10^{-2}$ and $dk_F \sim 1$, and thus the law of $\rho_{\rm d} \sim T^2$ may not be valid. On the other hand, we only have three temperatures of 200 K, 240 K and 300 K. More measurements at lower temperatures are required to offer better estimation of α. It unfortunate that the measurement of $\rho_{\rm d}$ for our samples below 200 K is challenging to perform and improvements are desired in future.

4.3 Counterflow thermoelectric transport in bilayer graphene double layer

High efficiency thermoelectric modules have been pursued for decades to potentially overcome the challenges regarding solid-state Peltier refrigerators and sustainable electricity harvested from the ubiquitous waste heat. Due to the conflicting material parameters, producing such modules requires careful material and structural engineering [144]. High efficiency in low dimensional and nanostructural engineered materials have been demonstrated, particularly for "electron-crystal-phonon-glass" systems that have high electrical conductivity and low thermal conductivity [144,145]. Many measurements show that graphene has high thermal conductivity κ ($\sim$ 600 – 3000 W/m-K) [51, 146], high electrical conductivity σ ($\sim 10^8/\Omega$-m) [147] and high Seebeck coefficient S ($\sim$100 μV/K) [148], resulting in the thermoelectric figure of merit $ZT = S^2\sigma T/\kappa \sim 0.01$ at room temperature. While small ZT materials are

not good candidates for thermoelectric applications, significant effort has been devoted to enhancing the ZT of graphene and related nanostructures [149–165], usually by reducing the thermal conductivity with various methods. These traditional approaches aim on engineering the properties of a single piece of material to enhance ZT. By bringing two pieces of material close at nanoscale, the Coulomb interaction between them can be strong, depending on the inter-piece distances. It remains primarily unknown how such interaction affects the thermoelectric transport properties. We thus study the alternative approach that employs the unique nanostructures of stacked layered systems with strong interlayer interactions, in particular to study the counterflow thermoelectric transport.

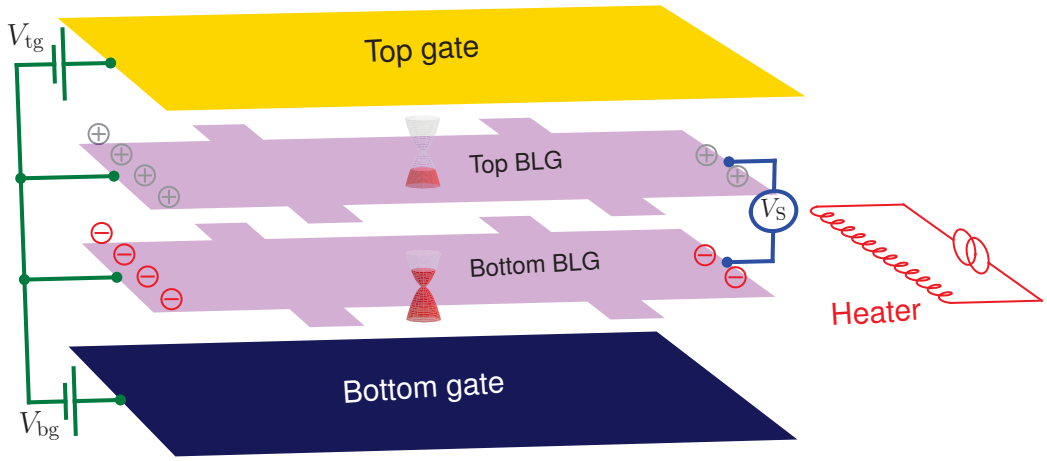

Fig. 4.10. Setup of electrical connection to measure the counterflow Seebeck coefficient.

The electrical connection to measure the counterflow Seebeck coefficient is shown in Fig. 4.10 where the counterflow Seebeck coefficient is specifically referenced to the voltage V_S measured between two layers at the right end while they are connected at

the left end when the temperature difference of 1 K is established between two ends. The top and bottom gate voltages $V_{\rm tg}$ and $V_{\rm bg}$ can control the carrier densities of the corresponding graphene layers, e.g., top layer p-type while the bottom layer n-type here indicated by the red filling of the parabolic band structures. The imbalanced carrier densities at two ends of each graphene layer driven by the temperature gradient are indicated by the symbols $\oplus$ for holes and $\ominus$ for electrons. A low frequency alternating heating current of a few milliampere is applied across the heater line and the counterflow Seebeck voltage $V_{\rm S}$ is then detected using a lock-in amplifier at the second harmonic. The heater and temperature sensors are calibrated [166] to convert the Seebeck voltage $V_{\rm S}$ to Seebeck coefficient $S_{\rm CF}$. The sign of $S_{\rm CF}$ is taken to be negative when the top and bottom graphene layers are p- and n-type respectively.

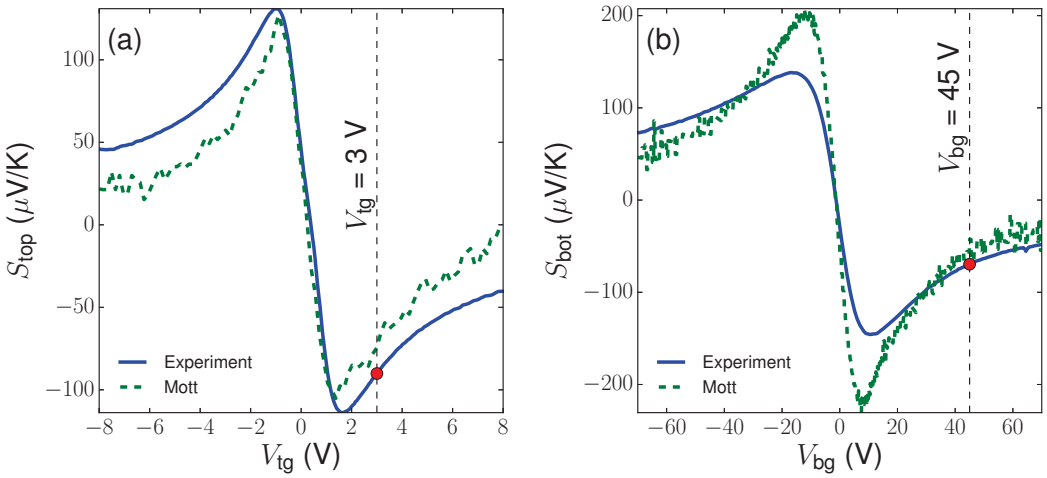

Fig. 4.11. Seebeck coefficient of top (a) and bottom (b) BLG layers (the green lines are calculated Seebeck coefficients from Mott formula) vs. $V_{\rm tg}$ and $V_{\rm bg}$ respectively.

We first present the measurement of the Seebeck coefficient for individual graphene layers, shown in Fig. 4.11. The Seebeck coefficient for both layers is positive (negative) when they are p-type (n-type), and is zero at the charge neutral point. The dashed lines represent the Seebeck coefficient calculated from the field effect curves in Fig. 4.4a and Fig. 4.4b, using the Mott formula

$$S = -\frac{\pi^2 k_B^2 T}{3eG}\frac{dG}{dV_g}\frac{dV_g}{dE_F} = \frac{\pi^2 k_B^2 T}{3e\rho}\frac{d\rho}{dV_g}\frac{dV_g}{dE_F} \tag{4.1}$$

where e is the elementary charge, k_B is the Boltzmann constant, V_g is the gate voltage, ρ (G) is the intralayer resistivity (conductivity), and E_F is the Fermi level. The $V_g \sim E_F$ relation can be obtained by evaluating the energy dispersion

$$E(k) = \pm\frac{1}{2}\gamma_1\left[\sqrt{1 + 4v_F^2\hbar^2k^2/\gamma_1^2} - 1\right] \tag{4.2}$$

(+ for conduction band and − for valence band) of the bilayer graphene at the Fermi wave vector $k_F = \sqrt{\pi|n|} = \sqrt{\pi C_g|V_g - V_D|/e}$, where n is the charge density, the Fermi velocity is $v_F = 10^6$ m/s, $\gamma_1 = 0.39$ eV, C_g is the gate capacitance and V_D is the gate voltage at the charge neutral point. An easier way is to evaluate dE_F/dV_g first and then do the inverse to obtain dV_g/dE_F. Note that the sign ± in Eq. (4.2) is canceled by the same sign from dk_F^2/dV_g. We arrive at

$$S = \frac{\pi k_B^2 T}{3\rho}\frac{d\rho}{dV_g}\frac{\gamma_1}{C_g v_F^2\hbar^2}\sqrt{1 + \frac{4v_F^2\hbar^2\pi C_g|V_g - V_D|}{\gamma_1^2 e}}. \tag{4.3}$$

Note that the wiggles in the green dashed lines are due to the numerical computation of $d\rho/dV_g$ from the lines in Fig. 4.4a and Fig. 4.4b.

The colormaps of counterflow Seebeck coefficient S_{CF} vs. V_{tg} and V_{bg} at four temperatures are shown in Fig. 4.12. Four quadrants of these colormaps are separated by white areas near zero gate voltages. Large magnitude of negative (positive) S_{CF} appears in the top left (bottom right) quadrant when the top and bottom graphene layers

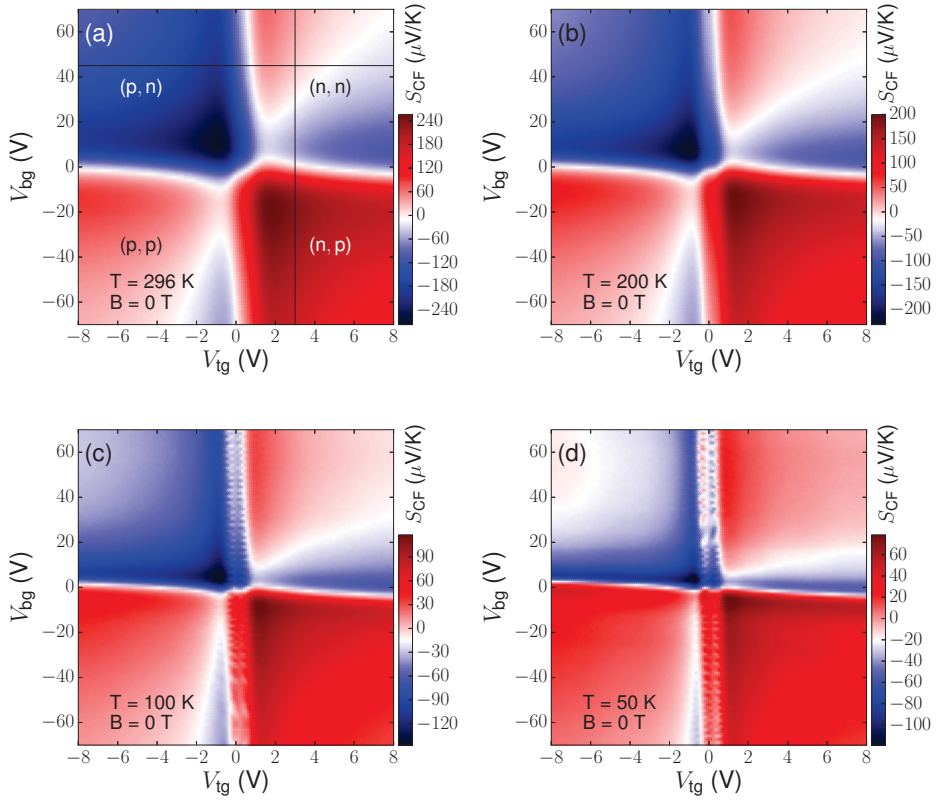

Fig. 4.12. Temperature dependent colormaps of counterflow Seebeck coefficient S_{CF} vs. V_{tg} and V_{bg}. The labels (x,y) in (a) indicate the x(y)-type of charge carriers for top (bottom) BLG.

are p- and n-type (n- and p-type) respectively. In this regime where two graphene layers have different carrier types, the magnitude of S_{CF} is simply the sum of the magnitude of Seebeck coefficient in each individual layer. The interlayer interaction that might affect it is neglected due to weak Coulomb drag in our samples (note ρ_d is nearly three orders of magnitude smaller than ρ_t or ρ_b). The magnitude of S_{CF} in the other two quadrants is smaller or even close to zero since the signs of Seebeck coefficient in both layers are same and they tend to cancel each other during the

counterflow measurement. The above discussion can be represented by the formula $S_{\mathrm{CF}} = S_{\mathrm{bot}} - S_{\mathrm{top}}$ which is demonstrated in Fig. 4.13. For example in Fig. 4.13a, the solid blue line is from the horizontal cut in Fig. 4.12a while the dashed green line is calculated from $S_{\mathrm{bot}} - S_{\mathrm{top}}$ where S_{bot} is a constant value taken at the red point in Fig. 4.11b for $V_{\mathrm{bg}} = 45$V and S_{top} is the solid blue curve in Fig. 4.11a. These two lines are close to each other, validating the above formula. The small discrepancy may come from the incomplete gating screening of graphene layers during the counterflow measurement. The horizontal lines of zero S_{CF} in Fig. 4.13 separate them into red (positive S_{CF}) and cyan (negative S_{CF}) areas. Note that the lines of S_{CF} vs. gate voltage cross zero twice, which is not possible for an isolated single layer of bilayer graphene. This explains that the top right and bottom left quadrants in Fig. 4.12 are composed of regions with different sign of S_{CF} and the boundaries of white line segments with zero S_{CF}.

As the temperature decreases, S_{CF} decreases and the red area of positive S_{CF} increases for the top right and bottom left quadrants in the colormaps. The positive and negative peaks of S_{CF} are slightly shifted closer to zero gate voltages and become more concentrated as the temperature is reduced. Many puddles of S_{CF} appear mainly around V_{tg} =0 V for temperatures 100 K and 50 K. This could be due to the manifestation of the charge inhomogeneity in the top graphene layer at lower temperatures and the small misalignment of the two layers during sample fabrication, resulting in disturbed Seebeck coefficient of the individual graphene layer around its charge neutral point.

The colormaps of c.c. of S_{CF} vs. T and the intercept b of the linear fit of $S_{\mathrm{CF}} \sim a\mathrm{T}+b$ are shown in Fig. 4.14a and Fig. 4.14b respectively. The blank area in Fig. 4.14b corresponds to the gate voltages at which c.c. in Fig. 4.14a is less than 0.99. The magnitude of the intercept in Fig. 4.14b is close to zero for the majority area of gate

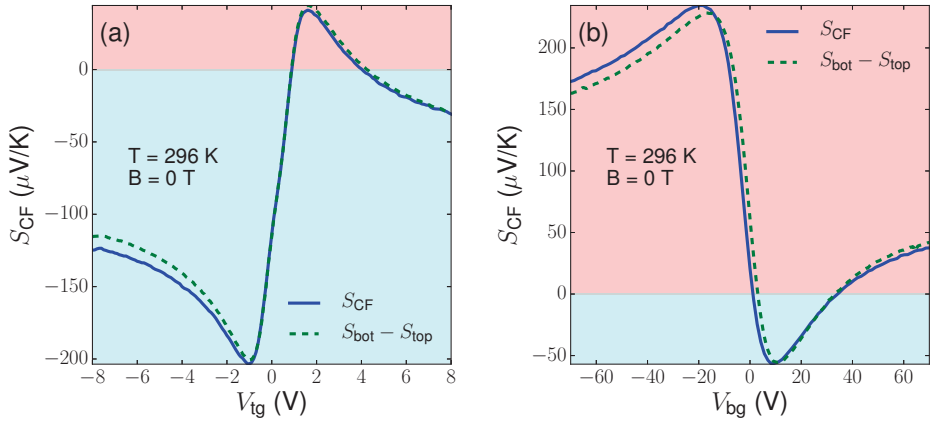

Fig. 4.13. S_{CF} vs. V_{tg} (a) for $V_{bg} = 45$ V and S_{CF} vs. V_{bg} for $V_{tg} = 3$ V. The solid blue line in (a(b)) is from the horizontal (vertical) cut in Fig. 4.12a. The dashed green line in (d(e)) is calculated from $S_{bot} - S_{top}$ where S_{bot} is a constant value at the red dot (the solid blue line) in Fig. 4.11b while S_{top} is the solid blue line (a constant value at the red dot) in Fig. 4.11a. The red (cyan) area in (a) and (b) represents the positive (negative) value of S_{CF}.

voltages away from the charge neutral point, indicating that S_{CF} is linearly dependent on temperature.

During the measurement of Seebeck voltage V_S shown in Fig. 4.10, the resistivities shown in Fig. 4.4 and Fig. 4.5 of both graphene layers are simultaneously measured using another two lock-in amplifiers at different lock-in frequencies. The frequencies for those three lock-in amplifiers are varied to ensure that the measured V_S and resistivities are immune to such variations, indicating a coexistence of independent thermoelectric and resistance response signals in each graphene layer.

The power factor in the counterflow thermoelectric transport regime is defined as $P = S_{CF}^2/[(\rho_t + \rho_b)t]$ where t=0.67 nm is the thickness of BLG. Its colormaps at four different temperatures are shown in Fig. 4.15. Note that the gate voltages

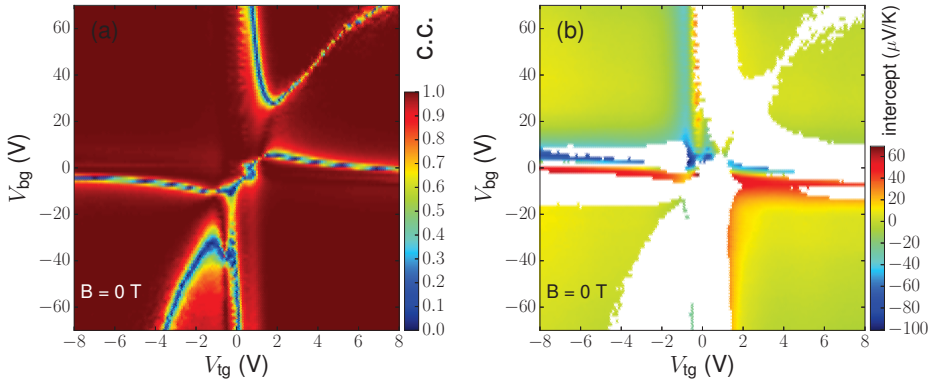

Fig. 4.14. Collormap (a) of correlation coefficient (c.c.) between S_{CF} and temperature and colormap (b) of the intercept b of the linear fit of $S_{\mathrm{CF}} \sim aT + b$. The blank area in (b) corresponds to the gate voltages at which c.c. in (a) is less than 0.99.

corresponding to the maximum of P are different from that of the magnitude of S_{CF} since the graphene resistivity is strongly dependent on the gate voltages. Here the extrinsic resistance mainly from the metal-graphene contacts is ignored, because it is material and process dependent and can be one order smaller than graphene resistance after special contact treatment [167]. The maximum power factor is $P_{\mathrm{max}} \sim 700$ $\mu\mathrm{W/K^2cm}$, higher than that of good thermoelectric materials such as Bi_2Te_3 [168].

The positive and negative peak values of S_{CF} vs. temperature are shown in Fig. 4.16a. The magnitudes of both peaks decrease with temperature. The maximum power factor P_{max} in the two regimes of opposite carrier types for the two graphene layers (solid red (dashed blue) line for top layer n-type (p-type) and bottom layer p-type (n-type)) vs. temperature is shown in Fig. 4.16b. These two lines show similar trend and manifest the symmetry between the two regimes. It is expected that P_{max} will increase as temperature increases above room temperature.

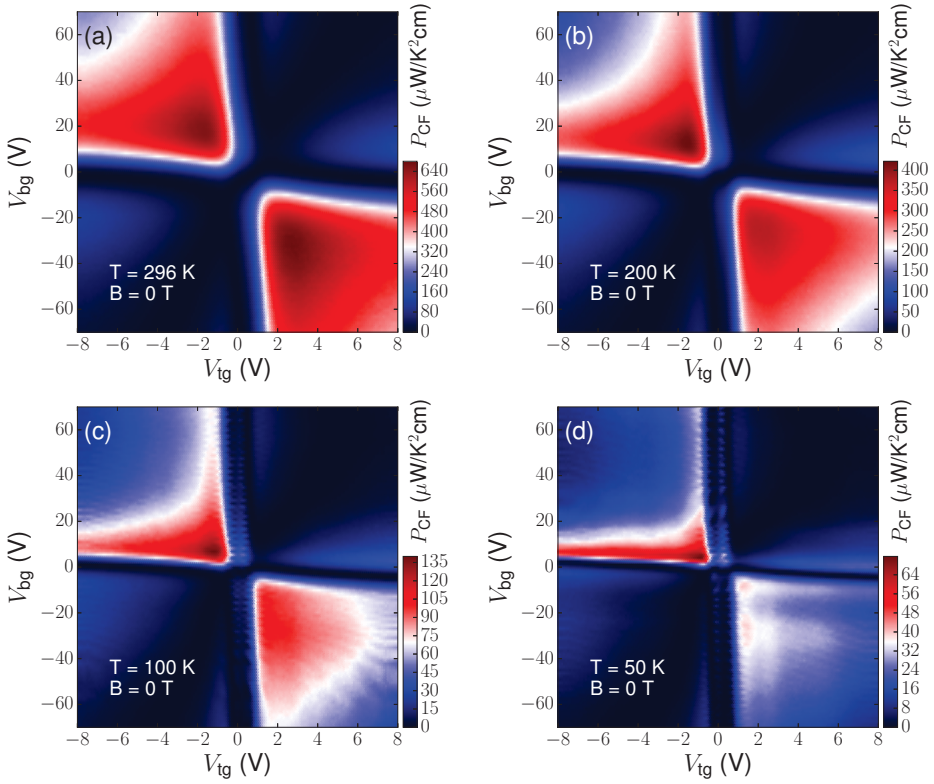

Fig. 4.15. Temperature dependent colormaps of counterflow power factor P_{CF} vs. V_{tg} and V_{bg}.

As we have pointed out that ZT of graphene is negligibly small, the high value of power factor is still promising in electricity generation at special circumstances, e.g., when space is constrained and high efficiency is not desired. If N_T is the number of thermoelectric units of Fig. 4.10c that are connected in series, the maximum output power is $P_{\mathrm{m}} = P_{\mathrm{max}} N_{\square} N_T t \Delta T^2/4$ at the impedance match, where $N_{\square}$ is the ratio of the width W to the length L (the length of graphene is measured along the direction of temperature gradient) and $P_{\mathrm{max}} \sim 700$ μW/K^{2}cm. For $\Delta T = 50$ K (e.g., the average temperature difference between the hot and cold spots in the modern CPUs),

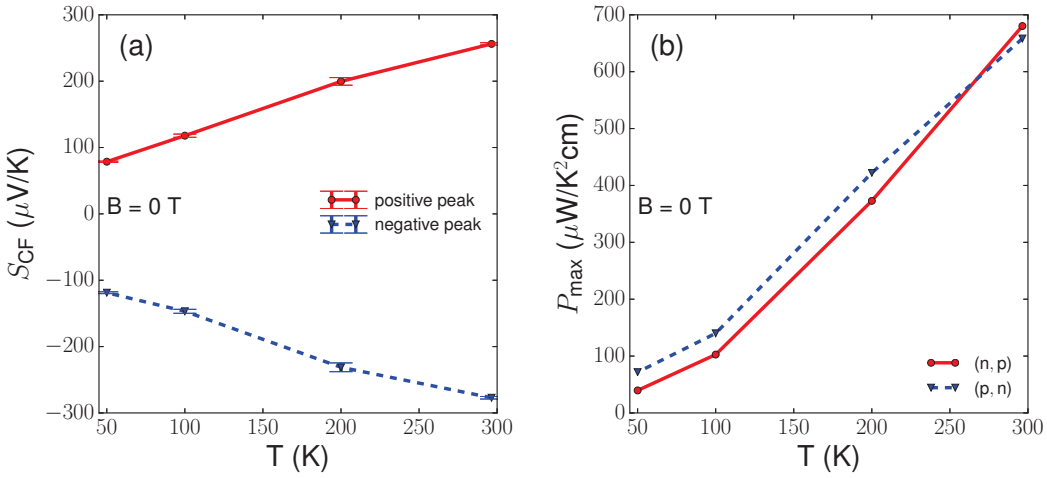

Fig. 4.16. The positive and negative peaks of $S_{\rm CF}$ (a) vs. temperature and the maximum power factor (b) $P_{\rm max}$ vs. temperature. The errorbars in (a) mainly from the calibration procedure for the temperature sensors represent the typical measurement error of $S_{\rm CF}$.

$P_{\rm m} = 2.9$ W for $N_\square = N_T = 10000$. Such serially connected thermopower generator has a thermal conductance of $G = 2\kappa N_T t N_\square = 201$ W/K for $\kappa = 1500$ W/m-K (G will be larger if the heat conduction channels through the insulating materials between graphene layers and the other supporting/gating materials are included), useful for applications of fast heat dissipation while generating electricity energy. Due to the low mass density of graphene, the power output per mass is as high as $\sim 1.1 \times 10^9$ W/kg (the total graphene area is ~ 20 μm^2) without including the mass of the insulating materials between graphene layers. It is still an extremely large number even when the mass of those insulating materials are included. For example, inserting a 3 nm thick of BN layer between graphene layers is good enough to avoid leakage between graphene layers. The mass density of BN is about 4.5 times larger than that of graphene, thus the power output per mass is reduced to $\sim 2.4 \times 10^8$

W/kg. The mass of the other materials such as the gate metal (can be removed by doping graphene chemically for example) and interconnects (could be replaced by graphene) is not considered here. Such high power per mass enables graphene in light weight thermoelectric applications [169].

4.4 Summary

Coulomb drag and counterflow thermoelectric transport measurements are performed in layered structures of BN/BLG/BN/BLG/BN. The magnitude of the counterflow Seebeck coefficient exhibits a peak in the regime where two graphene layers have opposite sign of charge carriers. The maximum power factor is about $700\mu W/K^2cm$ at room temperature. A quantitative analysis indicates that graphene can be useful for light weight thermoelectric systems. The counterflow Seebeck coefficient and power factor decrease approximately linearly with temperature from 300 K to 50 K. The measured interlayer Coulomb drag resistivity is negligibly small ($<$ 3 Ω), compared to the intralayer resistivity, suggesting negligible impact of Coulomb drag on the counterflow thermoelectric transport. Nevertheless, high Coulomb drag resistivity (esp. comparable to the intralayer resistivities) could strongly alter the properties of counterflow transport, which is the major subject of the next chapter.

5. TRANSPORT THEORY IN MULTILAYER SYSTEMS

For double layer system, the presence of weak Coulomb drag effect will not appreciably affect the (counterflow) thermoelectric transport. As seen from Fig. 4.9 and Fig. 4.8, the maximum magnitude of the Coulomb drag resistivity $\rho_{\rm d}$ is less than 10 Ω, much smaller than the resistivity of each graphene layer. However, it may become important when $\rho_{\rm d}$ is large, especially when it is close to the intralayer resistivities, which is the major topic of this chapter. Large Coulomb drag resistivity is an indication of strong interlayer interactions. For example, the formation of excitons composed of electron-hole pairs where the electrons and holes reside in different layers will possibly lead to high ZT structures [170, 171]. Although this chapter is focusing on a generic theory that is not specifically developed for graphene-based materials or structures, it is indeed directed from the experimental study of graphene double layer systems presented in the last chapter.

5.1 Coulomb drag and counterflow thermoelectric transport in double layer systems

It is unknown whether the magnitude of Coulomb drag resistivity can be comparable to or even larger than the intralayer resistivities. We now provide a phenomenological argument to derive a general relation between them in the following. We furthermore show that the existence of large Coulomb drag resistivity can dra-

matically enhance the figure of merit ZT in the counterflow thermoelectric transport regime.

For an isolated layer, the current flowing in it can be determined by the applied voltage at its two ends or the external electrical field inside and the resistivity of the layer. When two layers are coupled through the Coulomb drag resistivity, the current $I_{\rm b}$ flow in the bottom layer for example, will induce an electromotive force of $I_{\rm b}\rho_{\rm tb}\frac{L}{W}$ on the top layer, represented by the red voltage source symbol in Fig. 5.1. Here $\rho_{\rm tb}$ is the Coulomb drag resistivity when current is flowing in the bottom layer, L and W are the length (along current direction) and width (transverse to current direction) of each layer respectively. The top layer is then equivalent to a battery with an electromotive force of $I_{\rm b}\rho_{\rm tb}\frac{L}{W}$ and an internal resistance of $R_{\rm t}$ that is the top layer resistance. Now the top layer is connected in serial with a big resistor $R_{\rm o}$ and an ideal voltage source with electromotive force of $V_{\rm o}$ with the polarity indicated by the $+/-$ symbol inside the green circle. We assume that $V_{\rm o}$ is much larger than the magnitude of $I_{\rm b}\rho_{\rm tb}\frac{L}{W}$, so that the direction of the current flow in the top layer is solely determined by the polarity of the green voltage source. We also assume that $R_{\rm o}$ is much larger than $R_{\rm t}$ (we include any other parasitic resistances such as the contact resistance in $R_{\rm o}$). According to elementary circuits theory, the current flow in the top layer is

$$I_{\rm t} = \frac{V_{\rm o} - I_{\rm b}\rho_{\rm tb}\dfrac{L}{W}}{R_{\rm o} + R_{\rm t}} \approx \frac{V_{\rm o}}{R_{\rm o}}. \tag{5.1}$$

Thus, the prescribed current $I_{\rm t}$ can have any sign and magnitude, irrespective of the current in the bottom layer.

The output power of the green voltage source is $I_{\rm t}V_{\rm o}$. Then the electrical power transfered into the top layer is

$$P_{\rm t} = I_{\rm t}V_{\rm o} - I_{\rm t}^2 R_{\rm o} = I_{\rm t}\left(I_{\rm b}\rho_{\rm tb}\frac{L}{W} + I_{\rm t}R_{\rm t}\right) = I_{\rm t}\left(I_{\rm b}\rho_{\rm tb} + I_{\rm t}\rho_{\rm t}\right)\frac{L}{W} \equiv I_{\rm t}V_{\rm t}. \tag{5.2}$$

We can identify the voltage $V_{\rm t}$ as the total voltage drop, a sum of the voltage produced by the current flow in the layer and the Coulomb drag voltage due to the current flow in the other layer. Such decomposition process is illustrated in Fig. 5.2, where $V_{\rm t} = (\rho_{\rm t}I_{\rm t} + \rho_{\rm tb}I_{\rm b})\frac{L}{W}$ and $V_{\rm b} = (\rho_{\rm b}I_{\rm b} + \rho_{\rm bt}I_{\rm t})\frac{L}{W}$, which is valid in the linear response regime for small $I_{\rm t}$ and $I_{\rm b}$ that can be tuned in the way in Fig. 5.1. Here we consider the possibility that the Coulomb drag resistivity violates the Onsager's relation [172], i.e., $\rho_{\rm bt}$ and $\rho_{\rm tb}$ can be different, where $\rho_{\rm xy}$ is the drag resistivity when the current flows in the y layer while the voltage is measured in the x layer.

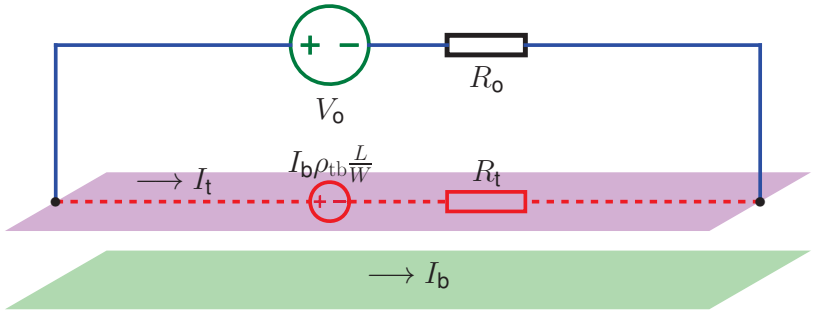

Fig. 5.1. Prescribing a current $I_{\rm t}$ on the top layer of the bilayer system when the bottom layer has current $I_{\rm b}$. The arrows $\longrightarrow$ represent the current direction.

Note that $I_{\rm t}$ and $V_{\rm t}$ (or similarly for $I_{\rm b}$ and $V_{\rm b}$) can have opposite signs, since now the sign of $I_{\rm t}$ is controlled by the external voltage source while the sign of $V_{\rm t}$ can be changed by the independent current $I_{\rm b}$.

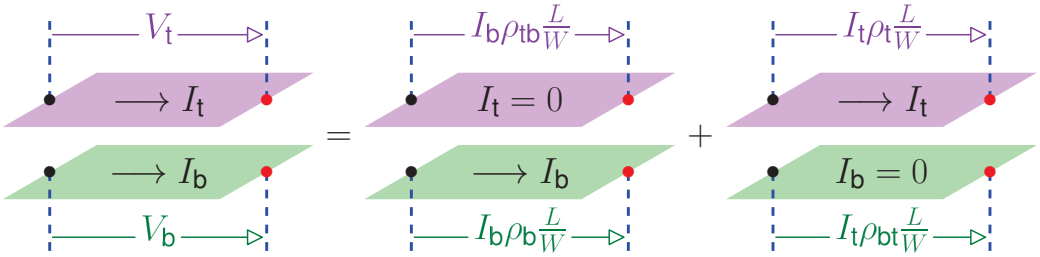

Fig. 5.2. Decomposition of voltage drop in each layer for generic double layer system with interlayer Coulomb drag. The formulas on arrows $\multimap$ represent the voltage drop along the arrows between two ends of them. The arrows $\longrightarrow$ represent the current direction. L and W are the length (along current direction) and width (transverse to current direction) of each layer respectively.

The total power $P = I_t V_t + I_b V_b$ consumed by the two layers cannot be negative, because there is no active electrical elements in the double layer system. Thus we must have

$$I_t^2 \rho_t + I_b^2 \rho_b + I_t I_b (\rho_{tb} + \rho_{bt}) \geq 0. \tag{5.3}$$

Since I_t and I_b can be arbitrarily small with arbitrary signs, we arrive at

$$-\sqrt{\rho_t \rho_b} \leq \frac{\rho_{tb} + \rho_{bt}}{2} \leq \sqrt{\rho_t \rho_b}. \tag{5.4}$$

If the Onsager relation holds, i.e., $\rho_{tb} = \rho_{bt} \equiv \rho_D$, we have $|\rho_D| \leq \sqrt{\rho_t \rho_b}$. Therefore, the magnitude of the Coulomb drag resistivity is limited by the intralayer resistivity. This is equivalent to require that the resistivity matrix

$$\rho = \begin{bmatrix} \rho_t & \rho_{tb} \\ \rho_{bt} & \rho_b \end{bmatrix} \tag{5.5}$$

is positive semi-definite,[1] since the total power consumed can be written in a quadratic form of $P = \mathbf{I}^T \rho \mathbf{I}$, where $\mathbf{I} = [I_\mathrm{t}\ I_\mathrm{b}]^T$ so that the voltage $\mathbf{V} = [V_\mathrm{t}\ V_\mathrm{b}]^T = \rho\mathbf{I}$, with the superscript ^{T} for vector transpose.

If the bottom layer is in a superconducting state, $\rho_\mathrm{b} = \rho_\mathrm{bt} = 0$ since no voltage drop can be established in the superconducting layer. Then obviously $\rho_\mathrm{tb} = 0$. This is a counterintuitive result, because the Cooper pairs in superconductors are charged and supposed to be able to drag the charge carriers in the adjacent normal conductor. Nevertheless, our prediction is consistent with the reported experiments [173–176].

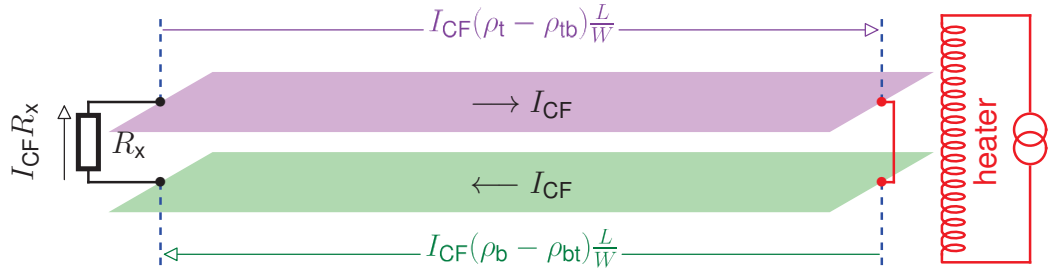

Fig. 5.3. Counterflow thermoelectric transport for generic double layer system with interlayer interaction. The formulas on arrows $\multimap\!\triangleright$ represent the voltage drop along the arrows between two ends of them. The arrows $\longrightarrow$ represent the current direction. L and W are the length (along current direction) and width (transverse to current direction) of each layer respectively.

[1] If ρ is positive semi-definite, a class of matrices

$$\begin{bmatrix} \rho_\mathrm{t} & \frac{\rho_\mathrm{tb}+\rho_\mathrm{bt}}{2} + \epsilon \\ \frac{\rho_\mathrm{tb}+\rho_\mathrm{bt}}{2} - \epsilon & \rho_\mathrm{b} \end{bmatrix}$$

for any $\epsilon \in \mathbb{R}$ are positive semi-definite. However they may not have real eigenvalues (e.g., when $|\epsilon|$ is large). Note that the Sylvester-like criteria for positive semi-definiteness (for 2×2 matrices, it requires non-negative diagonal entries and determinant) only applies for symmetrical matrices, e.g., $\epsilon = 0$.

The general counterflow thermoelectric transport can be represented by the schematic in Fig. 5.3. The two layers are shorted at the right side, and a resistance $R_{\rm x}$ is connected between two layers at the left side. Here $R_{\rm x}$ includes a load resistor $R_{\rm L}$ and the extrinsic resistances of contacts and interconnects. Assume that the counterflow Seebeck coefficient is $S_{\rm CF}$, the voltage measured between two red dots in Fig. 5.3 after the red wire connecting them is removed when a temperature difference of 1 K is established between the black dots and red dots. A heater at the right side will create a temperature difference ΔT between the two ends of each layer that induces an electromotive force $S_{\rm CF}\Delta T$ to drive the counterflow current $I_{\rm CF}$. From Fig. 5.3, it satisfies

$$I_{\rm CF}R_{\rm x} + I_{\rm CF}(\rho_{\rm t} - \rho_{\rm tb})\frac{L}{W} + I_{\rm CF}(\rho_{\rm b} - \rho_{\rm bt})\frac{L}{W} = S_{\rm CF}\Delta T. \tag{5.6}$$

We then have

$$I_{\rm CF} = \frac{S_{\rm CF}\Delta T}{\left[R_{\rm x} + (\rho_{\rm t} + \rho_{\rm b} - \rho_{\rm tb} - \rho_{\rm bt})\,\frac{L}{W}\right]}. \tag{5.7}$$

The effective electrical conductivity from Equation 5.7 is

$$\sigma_{\rm eff} = \frac{L}{t\left[WR_{\rm x} + (\rho_{\rm t} + \rho_{\rm b} - \rho_{\rm tb} - \rho_{\rm bt})\,L\right]}, \tag{5.8}$$

where t is the thickness of each layer. If the counterflow transport is dissipationless, i.e., $\rho_{\rm tb} + \rho_{\rm bt} = 2\rho_t = 2\rho_b$, $I_{\rm CF} = S_{\rm CF}\Delta T/R_{\rm x}$ and $\sigma_{\rm eff} = L/(tWR_{\rm x})$. Apparently, $ZT \propto L$ since $\sigma_{\rm eff}$ used for computing ZT is now proportional to the length of the thermoelectric unit. Thus, seeking bilayer systems with high Coulomb drag resistivity is desired to design novel high efficiency thermoelectric systems that can be simply optimized by the system geometry. Now the dissipation only happens at the load, contacts and interconnects. One obvious application is that it could promise dissipationless electrical power transmission. Note that if $\rho_{\rm tb} \neq \rho_{\rm bt}$, the power consumption in one layer is negative. This does not violate the law of energy conservation; energy is merely transferred between two layers via Coulomb interaction.

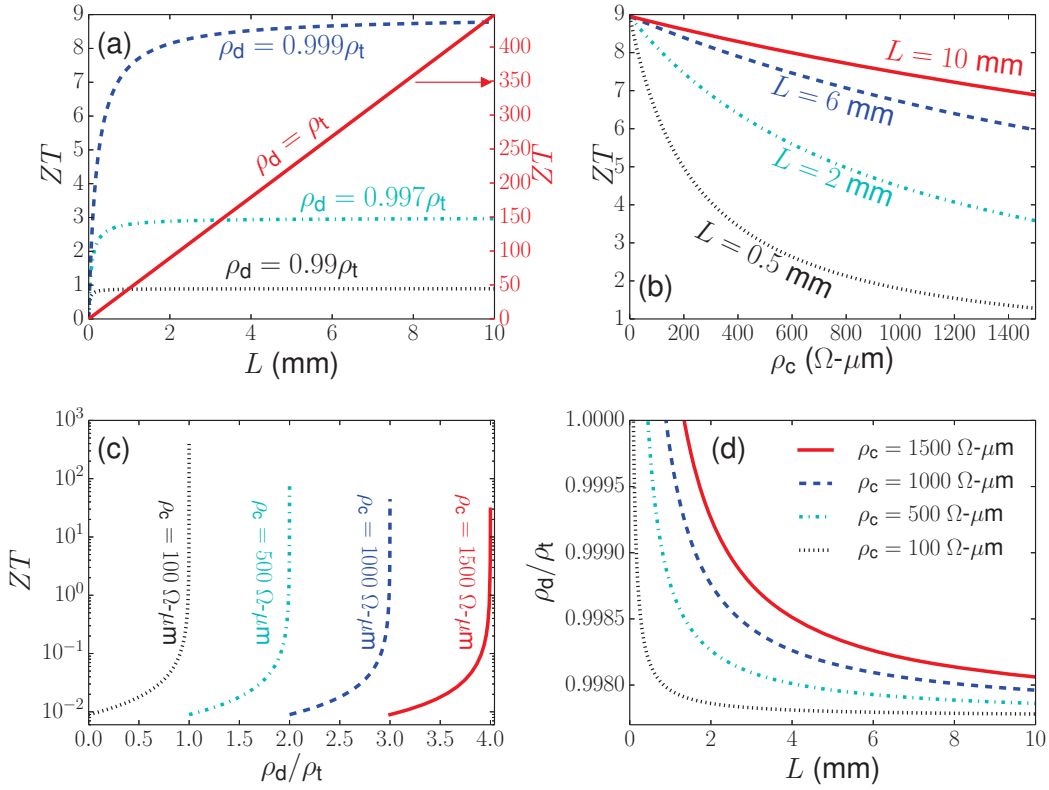

Fig. 5.4. (a) ZT vs. L at different values of ρ_d when $\rho_c = 100$ Ω-μm. Note that the vertical axis of the line of $\rho_d = \rho_t$ is on the right. (b) ZT vs. ρ_c at different values of L when $\rho_d = 0.999\rho_t$. (c) ZT vs. ρ_d/ρ_t at different value of ρ_c when $L = 10$ mm. All curves except the dotted one are shifted by integers horizontally for clarity. (d) ρ_d/ρ_t at $ZT = 4$ vs. L for different values of ρ_c.

To have positive $\rho_{bt(tb)}$, the two layers should have opposite sign of charge carriers. Graphene is then the ideal candidate to achieve $\rho_{tb} + \rho_{bt} = 2\rho_t = 2\rho_b$ for dissipationless counterflow transport due to its perfect electron-hole symmetry. Indeed, strong Coulomb drag has been observed in graphene/BN/graphene [130].

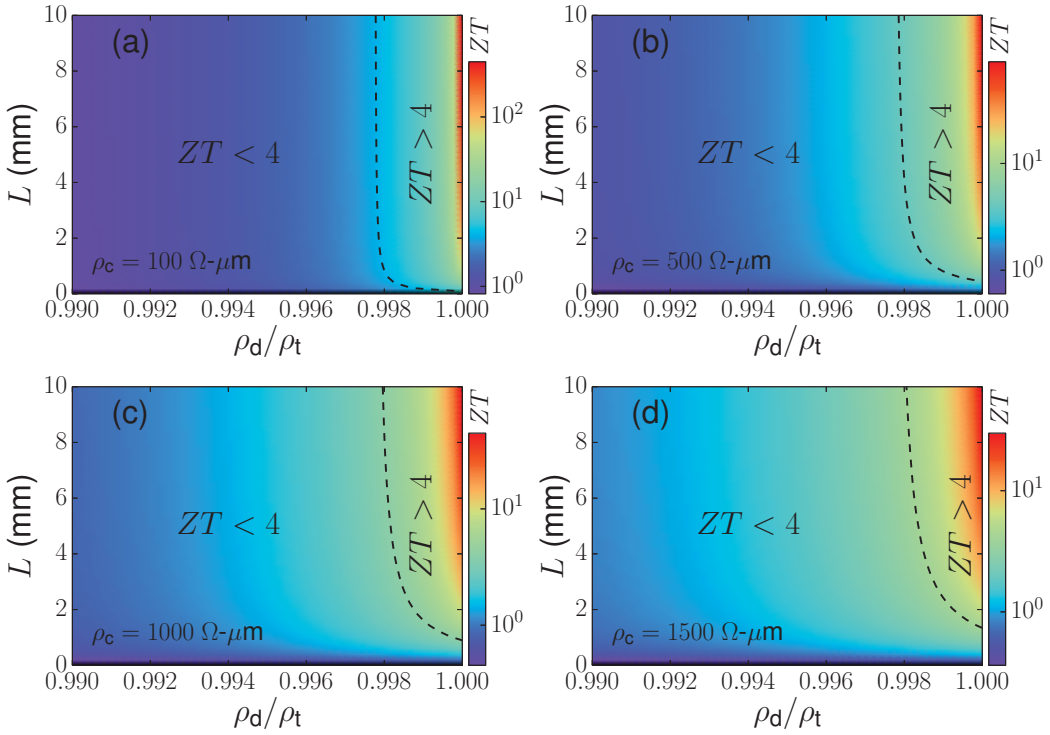

Fig. 5.5. Room temperature ZT vs. length L and normalized Coulomb drag resistivity ρ_d/ρ_t calculated from Equation 5.9 for four different values of ρ_c labeled in (a)-(d). $ZT = 4$ is indicated by the dashed lines above (below) which ZT is larger (smaller) than 4.

We now provide an explicit calculation of ZT for two layers of graphene in the counterflow transport regime. Graphene thickness is $t = 0.335$ nm. We assume that the intralayer resistivity [147] is $\rho_t = \rho_b = 1$ kΩ and the Coulomd drag resistivity is $\rho_{tb} = \rho_{bt} = \rho_d$. The counterflow Seebeck coefficient is assumed to be $S_{CF} = 200\ \mu$V/K [148]. The thermal conductivity of graphene at room temperature is taken as $\kappa = 2000$ W/m-K at room temperature [51,146]. We also assume that the extrinsic resistance is only from the 4 contacts, i.e., $R_x = 4\rho_c/W$ where ρ_c is the graphene

metal contact resistance ($100 - 1500$ Ω-μm) [167]. We neglect the contribution to the thermal conductivity from electrons since it is a few orders smaller than the lattice contribution [61]. Now ZT become

$$ZT = \frac{S_{\mathrm{CF}}^2 T}{\kappa t} \frac{L}{\left[4\rho_{\mathrm{c}} + 2\left(\rho_{\mathrm{t}} - \rho_{\mathrm{d}}\right) L\right]}. \tag{5.9}$$

The calculated ZT is shown in Fig. 5.5 for four different values of ρ_{c}. The white dashed lines represent $ZT = 4$ that the thermoelectric system starts to reach an efficiency comparable to commercial refrigerators [177]. The parameter space of L and ρ_d above these lines allows ZT to be larger than 4, which is significantly large for enabling large-scale thermoelectric applications.

5.2 Exciton and exciton condensation

In general, strong coupling between two layers is required to obtain high Coulomb drag resistivity. Intuitively, the formation of excitons will facilitate it. Under certain conditions if excitons are formed and the temperature is below a critical value that depends on the exciton density, the Bose-Einstein condensation of excitons will occur and thus exciton supercurrent will be supported (note that neutral excitons have repulsive interactions due to aligned electrical dipole moments). It is then desired to know the thermoelectric response of an exciton condensate. There are measurements of thermopower in Bose gases [178] and cuprate superconductors [179] where a negative sign of Seebeck coefficient is observed in the condensate regime. Similarly, the sign of the thermopower of exciton condensate should be negative if it is not zero (the negative sign means the exciton density is higher at heated spots). This can be understood by adopting the two fluid model with analogous to the superfluid fountain effect [180]. The conversion from superfluid component to normal component takes

place at the heated spots and results in superfluid flow towards the heated spots. Note that it is crucial to have a constriction (such as thin tubes or fine pores in observing the superfluid fountain effect) to slow down or stop the flow of the normal component in order to exhibit the fountain effect. As for exciton condensate, this requires that the sample bulk is resistive for the normal component of excitons, which results in a higher density of normal component of excitons at heated spots. This imbalance of normal component is the origin of generating the negative sign of thermopower of exciton condensate. One concern is that whether such imbalance of normal component can be maintained. There are two forces to reduce or even reverse this imbalance: diffusion and the Seebeck effect of excitons. It has been pointed out that, though the parallel transport in the regime of excitonic condensation is very resistive, the counterflow transport that supports exciton flow can be very conductive [181], and thus the imbalance is difficult to maintain. Note that the exciton condensate observed so far is in the bilayer quantum well systems with the same carrier types and densities for both layers in magnetic field at total filling fraction of $\nu = 1$, where the thermopower of excitons is zero due to the symmetry between two layers. The cannot be trivially generalized to the case with exciton condensate at zero magnetic field.

Even if an open circuit Seebeck voltage can be observed for an exciton condensate, this voltage can vanish when current flows. This is because the supercurrent and normal current components of the total electrical currents flowing in both layers will be adjusted to yield minimum Joule heat generation, resulting in equal electrical fields in both layers [182], and thus zero counterflow Seebeck coefficient. There are no experiments on measuring counterflow Seebeck coefficient for two layers with opposite types of carrier charge in the exciton condensation regime. It is still an open question whether the counterflow Seebeck coefficient is zero in the exciton condensation regime.

Nevertheless, if the pair formation of excitons happen at a temperature higher than the condensation temperature, which can be achieved at the BEC limit [183] where the size of the excitons is smaller than the inter-exciton distance, the counterflow Seebeck coefficient should be nonzero. Actually a recent work with a similar arguments states that the bilayer exciton can greatly enhance ZT [170].

5.3 Drude-like model for Coulomb drag

We offer a Drude-like model[2] to explore the possibility of dissipationless transport in double layer systems. To proceed with the spirit of Drude model, we firstly assume that the scattering event between the carrier in one layer and the ions in both layers is independent with the event between the carrier and the other carriers in the other layer. We can further assume that the carrier-carrier interaction within each layer is neglected, the intralayer momentum relaxation times are τ_{t} and τ_{b} (comes from the collisions between electrons and ions), and the momentum relaxation times due to the collisions between a pair of charge carriers residing in different layers are τ_{Dt} and τ_{Db} (τ_{Dt} refers to the mean free time of the carriers on the top layer due to the scattering by the carriers on the bottom layer). The carrier charge, density and the electrical field are denoted by q_χ, n_χ and $\mathbf{E}_\chi$ respectively in the top $\chi = \mathrm{t}$ and bottom $\chi = \mathrm{b}$ layers.

We assume all the scattering events happen locally. Within a macroscopically small area ΔS projected onto both layers along the surface norm (e.g., top layer) at some point, there are $N_\chi = n_\chi \Delta S$ carriers. Both N_{t} and N_{b} are large enough to carry out statistical treatment with enough accuracy[3]. The N_χ carriers are labeled

[2]The textbook by Ashcroft and Mermin [85] is referenced when this model is developed.

[3]If the system is homogeneous for both layers, ΔS can be taken as the area of both layers.

by a set of indices I_χ. Apparently the size of I_χ is $N_\chi \equiv |I_\chi|$. At time instant t, the momentum of each carriers is denoted by $\mathbf{p}_\chi^i(t)$ for $i \in I_\chi$. At time instant $t + \mathrm{d}t$, we can partition I_χ into three mutually disjoint subsets J_χ, K_χ and L_χ within which the carriers undergo no scattering, scattering by ions and interlayer Coulomb scatterings respectively. Apparently $|K_\chi|/N_\chi = \mathrm{d}t/\tau_\chi$ and $|L_\chi|/N_\chi = \mathrm{d}t/\tau_{\mathrm{D}\chi}$. After scattering, the momentum is denoted by $\mathbf{q}_\chi^i(t + \mathrm{d}t)$ for $i \in K_\chi \cup L_\chi$. For $i \in J_\chi$, $\mathbf{p}_\chi^i(t + \mathrm{d}t) = \mathbf{p}_\chi^i(t) + q_\chi \mathbf{E}_\chi \mathrm{d}t$. Now the total momentum at $t + \mathrm{d}t$ is

$$\sum_{i\in J_\chi} \mathbf{p}_\chi^i(t + \mathrm{d}t) + \sum_{i\in L_\chi} \mathbf{q}_\chi^i(t + \mathrm{d}t) = \sum_{i\in J_\chi} \mathbf{p}_\chi^i(t) + |J_\chi| q_\chi \mathbf{E}_\chi \mathrm{d}t + \sum_{i\in L_\chi} \mathbf{q}_\chi^i(t + \mathrm{d}t) \tag{5.10}$$

since $\sum_{i\in K_\chi} \mathbf{q}_\chi^i(t + \mathrm{d}t) = 0$.

Since the intralayer Coulomb scattering is binary, we have the restriction that $|L_\mathrm{t}| = |L_\mathrm{b}|$. For any given $i \in L_\mathrm{t}$, there is a partner $j \in L_\mathrm{b}$, so that the carrier i with momentum $\mathbf{p}_\mathrm{t}^i(t)$ in the top layer scatters with the carrier j with momentum $\mathbf{p}_\mathrm{b}^j(t)$ in the bottom layer, and ends up with momentums $\mathbf{q}_\mathrm{t}^i(t + \mathrm{d}t)$ and $\mathbf{q}_\mathrm{b}^j(t + \mathrm{d}t)$. In the following discussions, when such is are traversing L_t (L_b), the js are also traversing L_b (L_t). The momentum are conserved during the Coulomb scattering, i.e.,

$$\mathbf{p}_\mathrm{t}^i(t) + \mathbf{p}_\mathrm{b}^j(t) = \mathbf{q}_\mathrm{t}^i(t + \mathrm{d}t) + \mathbf{q}_\mathrm{b}^j(t + \mathrm{d}t). \tag{5.11}$$

Therefore, the sum of the total momentum for two layers now at time $t + \mathrm{d}t$ is

$$\sum_{\chi=\mathrm{t,b}} \left[\sum_{i\in J_\chi \cup L_\chi} \mathbf{p}_\chi^i(t) + |J_\chi| q_\chi \mathbf{E}_\chi \mathrm{d}t \right], \tag{5.12}$$

which is independent of the details of the interlayer Coulomb scattering, as expected.

We consider the following definition of average momentum and convert Eq. (5.10) into a equation about it, so that we need not to trace the detailed evolution of every carriers. The averaging about momentum at time t is defined as

$$\langle f\left(\mathbf{p}_\chi(t)\right)\rangle_A \equiv \frac{1}{|A|} \sum_{i\in A} f\left(\mathbf{p}_\chi^i(t)\right) = \langle f\left(\mathbf{p}_\chi(t)\right)\rangle_{I_\chi} \tag{5.13}$$

where $f(\cdot)$ is any function (e.g., kinetic energy) of momentum and $A \subset I_\chi$ is large enough to perform accurate statistical average. Note that we have assumed that at any time instant t the averaging over a subset (e.g., A) of the total N_χ carriers can faithfully represent the averaging over N_χ carriers. Now Eq. (5.10) becomes

$$\langle \mathbf{p}_\chi(t+\mathrm{d}t)\rangle_{I_\chi} = \left(1 - \frac{\mathrm{d}t}{\tau_\chi} - \frac{\mathrm{d}t}{\tau_{D\chi}}\right)[\langle \mathbf{p}_\chi(t)\rangle_{I_\chi} + q_\chi \mathbf{E}_\chi \mathrm{d}t] + \frac{\mathrm{d}t}{\tau_{D\chi}}\langle \mathbf{q}_\chi(t+\mathrm{d}t)\rangle_{L_\chi}. \quad (5.14)$$

Due to the momentum conservation in Eq. (5.11), we have

$$\sum_{\chi=\mathrm{t,b}} \langle \mathbf{q}_\chi(t+\mathrm{d}t)\rangle_{L_\chi} = \sum_{\chi=\mathrm{t,b}} \langle \mathbf{p}_\chi(t)\rangle_{L_\chi} = \sum_{\chi=\mathrm{t,b}} \langle \mathbf{p}_\chi(t)\rangle_{I_\chi}. \quad (5.15)$$

In the linear response regime, the imposed current in each layer should be very small, and the average momentums $\langle \mathbf{p}_\chi(t)\rangle_{I_\chi}$ should also be small since the current densities are proportional to the average momentums. We can thus linearize the scattered momentums $\langle \mathbf{q}_\chi(t+\mathrm{d}t)\rangle_{L_\chi}$ as

$$\begin{bmatrix} \langle \mathbf{q}_\mathrm{t}(t+\mathrm{d}t)\rangle_{L_\mathrm{t}} \\ \langle \mathbf{q}_\mathrm{b}(t+\mathrm{d}t)\rangle_{L_\mathrm{b}} \end{bmatrix} = \begin{bmatrix} \alpha & \beta \\ \bar{\alpha} & \bar{\beta} \end{bmatrix} \begin{bmatrix} \langle \mathbf{p}_\mathrm{t}(t)\rangle_{I_\mathrm{t}} \\ \langle \mathbf{p}_\mathrm{b}(t)\rangle_{I_\mathrm{b}} \end{bmatrix} \equiv \Gamma \begin{bmatrix} \langle \mathbf{p}_\mathrm{t}(t)\rangle_{I_\mathrm{t}} \\ \langle \mathbf{p}_\mathrm{b}(t)\rangle_{I_\mathrm{b}} \end{bmatrix}, \quad (5.16)$$

where $\alpha + \bar{\alpha} = 1$ and $\beta + \bar{\beta} = 1$. We will assume that the parameters α and β in the Γ matrix are time independent. This is because microscopically the Coulomb interaction potential is not explicitly time dependent, and so are those parameters describing the averaging of the scattering events. Now Eq. (5.14) becomes

$$\begin{aligned} \frac{\mathrm{d}\mathbf{p}_\mathrm{t}(t)}{\mathrm{d}t} &= q_\mathrm{t}\mathbf{E}_\mathrm{t} - \frac{1}{\tau_\mathrm{t}}\mathbf{p}_\mathrm{t}(t) + \frac{1}{\tau_\mathrm{Dt}}[-\bar{\alpha}\mathbf{p}_\mathrm{t}(t) + \beta\mathbf{p}_\mathrm{b}(t)], \\ \frac{\mathrm{d}\mathbf{p}_\mathrm{b}(t)}{\mathrm{d}t} &= q_\mathrm{b}\mathbf{E}_\mathrm{b} - \frac{1}{\tau_\mathrm{b}}\mathbf{p}_\mathrm{b}(t) + \frac{1}{\tau_\mathrm{Db}}[-\beta\mathbf{p}_\mathrm{b}(t) + \bar{\alpha}\mathbf{p}_\mathrm{t}(t)], \end{aligned} \quad (5.17)$$

where we have omitted the averaging brackets $\langle\cdot\rangle_{I_\chi}$. We can identify the average force on each carrier in the top layer as $\mathbf{F}_\mathrm{t} = \frac{1}{\tau_\mathrm{Dt}}[-\bar{\alpha}\mathbf{p}_\mathrm{t}(t) + \beta\mathbf{p}_\mathrm{b}(t)]$ due to the momentum transfer from the bottom layer and likewise the force on the carriers in

the bottom layer is $\mathbf{F}_\mathrm{b} = \frac{1}{\tau_\mathrm{Db}}[-\beta\mathbf{p}_\mathrm{b}(t) + \bar{\alpha}\mathbf{p}_\mathrm{t}(t)]$. Note that $\mathbf{F}_\mathrm{a}$ and $\mathbf{F}_\mathrm{b}$ are in the opposite direction but they in general have different magnitudes, since it is the total force due to momentum transfer, rather than the force per carrier, should have same magnitudes.

Note also that the following relation for the interlayer Coulomb scattering rates has to be satisfied,

$$\frac{\tau_\mathrm{Dt}}{\tau_\mathrm{Db}} = \frac{N_\mathrm{t}}{N_\mathrm{b}} = \frac{n_\mathrm{t}}{n_\mathrm{b}}. \tag{5.18}$$

5.3.1 Resistivity matrix

We assume the carrier masses are m_χ. The electrical current density now is $\mathbf{j}_\chi = q_\chi n_\chi \mathbf{p}_\chi(t)/m_\chi$. In the following we still keep $\bar{\alpha}$ and β as general parameters. The equations for the current density converted from Eq. 5.17 are

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix}\mathbf{j}_\mathrm{t}\\ \\ \mathbf{j}_\mathrm{b}\end{bmatrix} = \begin{bmatrix}\dfrac{q_\mathrm{t}^2 n_\mathrm{t}}{m_\mathrm{t}} & \\ & \dfrac{q_\mathrm{b}^2 n_\mathrm{b}}{m_\mathrm{b}}\end{bmatrix}\begin{bmatrix}\mathbf{E}_\mathrm{t}\\ \\ \mathbf{E}_\mathrm{b}\end{bmatrix} - \begin{bmatrix}\dfrac{1}{\tau_\mathrm{t}} + \dfrac{\bar{\alpha}}{\tau_\mathrm{Dt}} & -\dfrac{\beta}{\tau_\mathrm{Dt}}\dfrac{m_\mathrm{b}q_\mathrm{t}n_\mathrm{t}}{m_\mathrm{t}q_\mathrm{b}n_\mathrm{b}}\\ -\dfrac{\bar{\alpha}}{\tau_\mathrm{Db}}\dfrac{m_\mathrm{t}q_\mathrm{b}n_\mathrm{b}}{m_\mathrm{b}q_\mathrm{t}n_\mathrm{t}} & \dfrac{1}{\tau_\mathrm{b}} + \dfrac{\beta}{\tau_\mathrm{Db}}\end{bmatrix}\begin{bmatrix}\mathbf{j}_\mathrm{t}\\ \\ \mathbf{j}_\mathrm{b}\end{bmatrix} \tag{5.19}$$

and at stationary state we have

$$\begin{bmatrix}\dfrac{q_\mathrm{t}^2 n_\mathrm{t}}{m_\mathrm{t}} & \\ & \dfrac{q_\mathrm{b}^2 n_\mathrm{b}}{m_\mathrm{b}}\end{bmatrix}\begin{bmatrix}\mathbf{E}_\mathrm{t}\\ \\ \mathbf{E}_\mathrm{b}\end{bmatrix} = \begin{bmatrix}\dfrac{1}{\tau_\mathrm{t}} + \dfrac{\bar{\alpha}}{\tau_\mathrm{Dt}} & -\dfrac{\beta}{\tau_\mathrm{Dt}}\dfrac{m_\mathrm{b}q_\mathrm{t}n_\mathrm{t}}{m_\mathrm{t}q_\mathrm{b}n_\mathrm{b}}\\ -\dfrac{\bar{\alpha}}{\tau_\mathrm{Db}}\dfrac{m_\mathrm{t}q_\mathrm{b}n_\mathrm{b}}{m_\mathrm{b}q_\mathrm{t}n_\mathrm{t}} & \dfrac{1}{\tau_\mathrm{b}} + \dfrac{\beta}{\tau_\mathrm{Db}}\end{bmatrix}\begin{bmatrix}\mathbf{j}_\mathrm{t}\\ \\ \mathbf{j}_\mathrm{b}\end{bmatrix} \tag{5.20}$$

The resistivity matrix then is

$$\begin{aligned}\rho &= \begin{bmatrix} \dfrac{m_\mathrm{t}}{q_\mathrm{t}^2 n_\mathrm{t}} & \\ & \dfrac{m_\mathrm{b}}{q_\mathrm{b}^2 n_\mathrm{b}} \end{bmatrix} \begin{bmatrix} \dfrac{1}{\tau_\mathrm{t}} + \dfrac{\bar{\alpha}}{\tau_\mathrm{Dt}} & -\dfrac{\beta}{\tau_\mathrm{Dt}} \dfrac{m_\mathrm{b} q_\mathrm{t} n_\mathrm{t}}{m_\mathrm{t} q_\mathrm{b} n_\mathrm{b}} \\ -\dfrac{\bar{\alpha}}{\tau_\mathrm{Db}} \dfrac{m_\mathrm{t} q_\mathrm{b} n_\mathrm{b}}{m_\mathrm{b} q_\mathrm{t} n_\mathrm{t}} & \dfrac{1}{\tau_\mathrm{b}} + \dfrac{\beta}{\tau_\mathrm{Db}} \end{bmatrix} \\ &= \begin{bmatrix} \dfrac{m_\mathrm{t}}{q_\mathrm{t}^2 n_\mathrm{t}} \left(\dfrac{1}{\tau_\mathrm{t}} + \dfrac{\bar{\alpha}}{\tau_\mathrm{Dt}} \right) & -\dfrac{\beta}{\tau_\mathrm{Dt}} \dfrac{m_\mathrm{b}}{n_\mathrm{b}} \dfrac{1}{q_\mathrm{t} q_\mathrm{b}} \\ -\dfrac{\bar{\alpha}}{\tau_\mathrm{Db}} \dfrac{m_\mathrm{t}}{n_\mathrm{t}} \dfrac{1}{q_\mathrm{t} q_\mathrm{b}} & \dfrac{m_\mathrm{b}}{q_\mathrm{b}^2 n_\mathrm{b}} \left(\dfrac{1}{\tau_\mathrm{b}} + \dfrac{\beta}{\tau_\mathrm{Db}} \right) \end{bmatrix}. \end{aligned} \tag{5.21}$$

We can identify

$$\begin{aligned} \rho_\mathrm{t} &= \frac{m_\mathrm{t}}{q_\mathrm{t}^2 n_\mathrm{t}} \left(\frac{1}{\tau_\mathrm{t}} + \frac{\bar{\alpha}}{\tau_\mathrm{Dt}} \right), \qquad \rho_\mathrm{bt} = -\frac{\bar{\alpha}}{\tau_\mathrm{Db}} \frac{m_\mathrm{t}}{n_\mathrm{t}} \frac{1}{q_\mathrm{t} q_\mathrm{b}}, \\ \rho_\mathrm{b} &= \frac{m_\mathrm{b}}{q_\mathrm{b}^2 n_\mathrm{b}} \left(\frac{1}{\tau_\mathrm{b}} + \frac{\beta}{\tau_\mathrm{Db}} \right), \qquad \rho_\mathrm{tb} = -\frac{\beta}{\tau_\mathrm{Dt}} \frac{m_\mathrm{b}}{n_\mathrm{b}} \frac{1}{q_\mathrm{t} q_\mathrm{b}}. \end{aligned} \tag{5.22}$$

Since the system is not in a magnetic field, the external electrical fields $\mathbf{E}_\chi$ are Galilean invariant. It seems that the friction term $-\frac{1}{\tau_\chi}\mathbf{p}_\chi(t)$ in Eq. (5.17) breaks the Galilean invariance, however, it is actually reduced from

$$-\frac{1}{\tau_\chi}\mathbf{p}_\chi(t) = -\frac{1}{\tau_\chi}\left[\langle \mathbf{p}_\chi(t)\rangle_{K_\chi} - \langle \mathbf{q}_\chi(t+\mathrm{d}t)\rangle_{K_\chi}\right], \tag{5.23}$$

which is Galilean invariant. The forces F_χ should be Galilean invariant, i.e., given any constant velocity $\mathbf{v}$, we should have

$$-\bar{\alpha}\left[\mathbf{p}_\mathrm{t}(t) + m_\mathrm{t}\mathbf{v}\right] + \beta\left[\mathbf{p}_\mathrm{b}(t) + m_\mathrm{b}\mathbf{v}\right] = -\bar{\alpha}\mathbf{p}_\mathrm{t}(t) + \beta\mathbf{p}_\mathrm{b}(t), \tag{5.24}$$

and we arrive at

$$\bar{\alpha} m_\mathrm{t} = \beta m_\mathrm{b}. \tag{5.25}$$

Therefore, the resistivity matrix is symmetrical, i.e.,

$$\rho_\mathrm{tb} = \rho_\mathrm{bt} \tag{5.26}$$

due to the relations in Eq. (5.18) and Eq. (5.25). Thus, it is the Galilean invariance of the double layer system that guarantees the Onsager relation in Eq. (5.26) for the Coulomb drag resistivities.

5.3.2 Elastic Coulomb scattering

We will figure out the values of the parameters $\bar{\alpha}$ and β for the following specific case. The kinetic energy conversation corresponding to Eq. (5.11) is:

$$\frac{[\mathbf{p}_{\mathrm{t}}^{i}(t)]^2}{2m_{\mathrm{t}}}+\frac{[\mathbf{p}_{\mathrm{b}}^{j}(t)]^2}{2m_{\mathrm{b}}}=\frac{[\mathbf{q}_{\mathrm{t}}^{i}(t+\mathrm{d}t)]^2}{2m_{\mathrm{t}}}+\frac{[\mathbf{q}_{\mathrm{b}}^{j}(t+\mathrm{d}t)]^2}{2m_{\mathrm{b}}}. \tag{5.27}$$

For convenience, we drop the time brackets in Eq. (5.11) and Eq. (5.27) for now. We can introduce a vector $\mathbf{k}_{ij}$, such that $\mathbf{q}_{\mathrm{t}}^{i}=\mathbf{p}_{\mathrm{t}}^{i}+\mathbf{k}_{ij}$ and $\mathbf{q}_{\mathrm{b}}^{j}=\mathbf{p}_{\mathrm{b}}^{j}-\mathbf{k}_{ij}$. Eq. (5.27) leads to

$$\mathbf{k}_{ij}^2\left(\frac{1}{2m_{\mathrm{t}}}+\frac{1}{2m_{\mathrm{b}}}\right)+\mathbf{k}_{ij}\cdot\mathbf{v}_{ij}=0, \tag{5.28}$$

where $\mathbf{v}_{ij}\equiv\frac{\mathbf{p}_{\mathrm{t}}^{i}}{m_{\mathrm{t}}}-\frac{\mathbf{p}_{\mathrm{b}}^{j}}{m_{\mathrm{b}}}$ is the relative velocity between two carriers before scattering. Assume the angle between $\mathbf{k}_{ij}$ and $\mathbf{v}_{ij}$ is θ_{ij}, then

$$\mathbf{k}_{ij}=-2\mu\mathbf{v}_{ij}\cos^2\theta_{ij}+\mu|\mathbf{v}_{ij}|\hat{\mathbf{v}}_{ij}^{\perp}\sin(2\theta_{ij}), \tag{5.29}$$

where $\mu\equiv\frac{m_{\mathrm{t}}m_{\mathrm{b}}}{m_{\mathrm{t}}+m_{\mathrm{b}}}$, $\hat{\mathbf{v}}_{ij}^{\perp}$ is a unit vector perpendicular to $\mathbf{v}_{ij}$, and θ can be any value in the range $[\pi/2,3\pi/2]$.

When performing the average over $\mathbf{k}_{ij}$ like $\sum_{i\in L_{\mathrm{t}}}\mathbf{k}_{ij}$, the second term on the right hand side in Eq. (5.29) yields zero since $\sin(2\theta_{ij})$ averages to zero for $\theta_{ij}\in[\pi/2,3\pi/2]$. However due to

$$\frac{1}{\pi}\int_{\pi/2}^{3\pi/2}\cos^2\theta\mathrm{d}\theta=\frac{1}{2}, \tag{5.30}$$

we have

$$\begin{aligned}
\langle \mathbf{q}_{\mathrm{t}}(t+\mathrm{d}t)\rangle_{L_{\mathrm{t}}} &= \frac{1}{|L_{\mathrm{t}}|}\sum_{i\in L_{\mathrm{t}}}(\mathbf{p}_{\mathrm{t}}^{i}+\mathbf{k}_{ij}) \\
&= \langle \mathbf{p}_{\mathrm{t}}(t)\rangle_{L_{\mathrm{t}}} - \frac{m_{\mathrm{t}}m_{\mathrm{b}}}{m_{\mathrm{t}}+m_{\mathrm{b}}}\left[\frac{\langle \mathbf{p}_{\mathrm{t}}(t)\rangle_{L_{\mathrm{t}}}}{m_{\mathrm{t}}} - \frac{\langle \mathbf{p}_{\mathrm{b}}(t)\rangle_{L_{\mathrm{b}}}}{m_{\mathrm{b}}}\right], \\
\langle \mathbf{q}_{\mathrm{b}}(t+\mathrm{d}t)\rangle_{L_{\mathrm{b}}} &= \frac{1}{|L_{\mathrm{b}}|}\sum_{j\in L_{\mathrm{b}}}(\mathbf{p}_{\mathrm{b}}^{j}-\mathbf{k}_{ij}) \\
&= \langle \mathbf{p}_{\mathrm{b}}(t)\rangle_{L_{\mathrm{b}}} + \frac{m_{\mathrm{t}}m_{\mathrm{b}}}{m_{\mathrm{t}}+m_{\mathrm{b}}}\left[\frac{\langle \mathbf{p}_{\mathrm{t}}(t)\rangle_{L_{\mathrm{t}}}}{m_{\mathrm{t}}} - \frac{\langle \mathbf{p}_{\mathrm{b}}(t)\rangle_{L_{\mathrm{b}}}}{m_{\mathrm{b}}}\right],
\end{aligned} \tag{5.31}$$

where in the first equation $j \in L_{\mathrm{b}}$ varies with i and likewise in the second equation. Thus, we can identify

$$\bar{\alpha} = \frac{m_{\mathrm{b}}}{m_{\mathrm{t}}+m_{\mathrm{b}}},\ \beta = \frac{m_{\mathrm{t}}}{m_{\mathrm{t}}+m_{\mathrm{b}}}. \tag{5.32}$$

5.3.3 Kinetic energy change

The change of the total kinetic energy E is $E(t+\mathrm{d}t) - E(t)$:

$$\begin{aligned}
&\sum_{\chi=\mathrm{t,b}}\left[\sum_{i\in J_\chi}\frac{\left(\mathbf{p}_\chi^i(t+\mathrm{d}t)\right)^2}{2m_\chi} + \sum_{i\in K_\chi\cup L_\chi}\frac{\left(\mathbf{q}_\chi^i(t+\mathrm{d}t)\right)^2}{2m_\chi} - \sum_{i\in I_\chi}\frac{\left(\mathbf{p}_\chi^i(t)\right)^2}{2m_\chi}\right] \\
&= \sum_{\chi=\mathrm{t,b}}\left[\sum_{i\in J_\chi}\frac{\left(\mathbf{p}_\chi^i(t+\mathrm{d}t)\right)^2 - \left(\mathbf{p}_\chi^i(t)\right)^2}{2m_\chi} + \sum_{i\in K_\chi}\frac{\left(\mathbf{q}_\chi^i(t+\mathrm{d}t)\right)^2 - \left(\mathbf{p}_\chi^i(t)\right)^2}{2m_\chi}\right] \\
&= \sum_{\chi=\mathrm{t,b}} N_\chi\left[q_\chi\mathbf{E}_\chi\cdot\frac{\mathbf{p}_\chi(t)}{m_\chi} - \frac{1}{\tau_\chi}\frac{\left(\mathbf{p}_\chi(t)\right)^2}{2m_\chi}\right]\mathrm{d}t,
\end{aligned} \tag{5.33}$$

since $\left\langle \left(\mathbf{p}_\chi(t)\right)^2\right\rangle_{K_\chi} = \left(\mathbf{p}_\chi(t)\right)^2 + 2m_\chi k_B T$ and $\left\langle \left(\mathbf{q}_\chi(t+\mathrm{d}t)\right)^2\right\rangle_{K_\chi} = 2m_\chi k_B T$ where k_B is the Boltzmann constant and T is the temperature. We see that at the steady state all the work done by the external fields is dissipated through the intralayer scattering. Note that the parameters $\tau_{\mathrm{D}\chi}$ is not explicitly present, since the interlayer Coulomb scattering is assumed to be elastic. This however don't imply that there is

no energy transfer between two layers. We can now calculate this transfer rate. For example, the energy transfer from the bottom layer to the top layer is:

$$\begin{aligned}
\Delta E_{\mathrm{b}\to\mathrm{t}} &\equiv \sum_{i\in L_{\mathrm{t}}} \frac{\left(\mathbf{q}_{\mathrm{t}}^{i}(t+\mathrm{d}t)\right)^2 - \left(\mathbf{p}_{\mathrm{t}}^{i}(t)\right)^2}{2m_{\mathrm{t}}} = \sum_{i\in L_{\mathrm{t}}} \frac{\left(\mathbf{p}_{\mathrm{t}}^{i}(t)+\mathbf{k}_{ij}\right)^2 - \left(\mathbf{p}_{\mathrm{t}}^{i}(t)\right)^2}{2m_{\mathrm{t}}} \\
&= \frac{1}{2m_{\mathrm{t}}} \sum_{i\in L_{\mathrm{t}}} \left[4\mu^2 \mathbf{v}_{ij}^2 \cos^2\theta_{ij} + 2\mathbf{p}_{\mathrm{t}}^{i}(t)\cdot\left(-2\mu\mathbf{v}_{ij}\cos^2\theta_{ij} + \mathbf{v}_{ij}^{\perp}\sin(2\theta_{ij})\right)\right] \\
&= \frac{\mu}{m_{\mathrm{t}}} \sum_{i\in L_{\mathrm{t}}} \left[\mu\mathbf{v}_{ij}^2 - \mathbf{p}_{\mathrm{t}}^{i}(t)\cdot\mathbf{v}_{ij}\right] \\
&= \frac{\mu}{m_{\mathrm{t}}+m_{\mathrm{b}}} \sum_{i\in L_{\mathrm{t}}} \left[\frac{\left(\mathbf{p}_{\mathrm{b}}^{j}(t)\right)^2}{m_{\mathrm{b}}} - \frac{\left(\mathbf{p}_{\mathrm{t}}^{i}(t)\right)^2}{m_{\mathrm{t}}} + \mathbf{p}_{\mathrm{t}}^{i}(t)\cdot\mathbf{p}_{\mathrm{b}}^{j}(t)\left(\frac{1}{m_{\mathrm{b}}} - \frac{1}{m_{\mathrm{t}}}\right)\right] \\
&= \frac{\mu}{m_{\mathrm{t}}+m_{\mathrm{b}}} \frac{N_{\mathrm{t}}\mathrm{d}t}{\tau_{\mathrm{D}\mathrm{t}}} \left[\frac{\left(\mathbf{p}_{\mathrm{b}}(t)\right)^2}{m_{\mathrm{b}}} - \frac{\left(\mathbf{p}_{\mathrm{t}}(t)\right)^2}{m_{\mathrm{t}}} + \mathbf{p}_{\mathrm{t}}(t)\cdot\mathbf{p}_{\mathrm{b}}(t)\left(\frac{1}{m_{\mathrm{b}}} - \frac{1}{m_{\mathrm{t}}}\right)\right] \\
&= \frac{N_{\mathrm{t}}\mathrm{d}t}{\tau_{\mathrm{D}\mathrm{t}}}\mu\left(\frac{\mathbf{p}_{\mathrm{b}}(t)}{m_{\mathrm{b}}} - \frac{\mathbf{p}_{\mathrm{t}}(t)}{m_{\mathrm{t}}}\right)\cdot\frac{\mathbf{p}_{\mathrm{t}}(t)+\mathbf{p}_{\mathrm{b}}(t)}{m_{\mathrm{t}}+m_{\mathrm{b}}} \\
&= N_{\mathrm{t}}\mathrm{d}t\mathbf{F}_{\mathrm{t}}\cdot\frac{\mathbf{p}_{\mathrm{t}}(t)+\mathbf{p}_{\mathrm{b}}(t)}{m_{\mathrm{t}}+m_{\mathrm{b}}}.
\end{aligned} \tag{5.34}$$

Similarly we have $\Delta E_{\mathrm{t}\to\mathrm{b}} = N_{\mathrm{b}}\mathrm{d}t\mathbf{F}_{\mathrm{b}}\cdot\frac{\mathbf{p}_{\mathrm{t}}(t)+\mathbf{p}_{\mathrm{b}}(t)}{m_{\mathrm{t}}+m_{\mathrm{b}}}$. We then have $\Delta E_{\mathrm{b}\to\mathrm{t}} + \Delta E_{\mathrm{t}\to\mathrm{b}} = 0$, as expected for elastic Coulomb scatterings. Naively, the kinetic energy change due to interlayer interaction would be $\Delta E_{\mathrm{b}\to\mathrm{t}} = N_{\mathrm{t}}\mathrm{d}t\mathbf{F}_{\mathrm{t}}\cdot\frac{\mathbf{p}_{\mathrm{t}}(t)}{m_{\mathrm{t}}}$ and $\Delta E_{\mathrm{t}\to\mathrm{b}} = N_{\mathrm{b}}\mathrm{d}t\mathbf{F}_{\mathrm{b}}\cdot\frac{\mathbf{p}_{\mathrm{b}}(t)}{m_{\mathrm{b}}}$. Then $\Delta E_{\mathrm{b}\to\mathrm{t}} + \Delta E_{\mathrm{t}\to\mathrm{b}} \neq 0$ in general. This is similar to the situation in Eq. (5.33) where the energy loss due to the intralayer collisions is $-\frac{1}{\tau_\chi}\frac{\left(\mathbf{p}_\chi(t)\right)^2}{2m_\chi}\mathrm{d}t$ which is not equal to the product of the friction force $-\frac{\mathbf{p}_\chi(t)}{\tau_\chi}$ and the average velocity $\frac{\mathbf{p}_\chi(t)}{m_\chi}$. We thus note that the friction and Coulomb drag forces ($\mathbf{F}_\chi$) are fundamentally different from the external forces, though all of these describe the change of the average momentum, as in Eq. (5.17). The former ones are converted from the average momentum change per carrier due to the scatterings where not all the carriers are involved during the time interval $\mathrm{d}t$. However, all of the carriers are accelerated/decelerated by the external fields.

5.3.4 Discussions on the Drude-like model

The determinant of the resistivity matrix is

$$\begin{aligned} |\rho| = \rho_{\rm t}\rho_{\rm b} - \rho_{\rm tb}\rho_{\rm bt} &= \frac{m_{\rm t}m_{\rm b}}{q_{\rm t}^2 q_{\rm b}^2 n_{\rm t} n_{\rm b}} \left(\frac{1}{\tau_{\rm t}\tau_{\rm b}} + \frac{\bar{\alpha}}{\tau_{\rm b}\tau_{\rm Dt}} + \frac{\beta}{\tau_{\rm t}\tau_{\rm Db}} \right) \\ &= \frac{m_{\rm t}}{q_{\rm t}^2 n_{\rm t}\tau_{\rm t}} \frac{m_{\rm b}}{q_{\rm b}^2 n_{\rm b}\tau_{\rm b}} \left[1 + \gamma\omega \left(\frac{\tau_{\rm t}}{m_{\rm t}n_{\rm t}} + \frac{\tau_{\rm b}}{m_{\rm b}n_{\rm b}} \right) \right], \end{aligned} \tag{5.35}$$

where we denote $\gamma \equiv \bar{\alpha} m_{\rm t} = \beta m_{\rm b}$ and $\omega \equiv \dfrac{n_{\rm t}}{\tau_{\rm Dt}} = \dfrac{n_{\rm b}}{\tau_{\rm Db}}$. Then the positive semi-definiteness of ρ is equivalent to

$$\gamma\omega \geq -\left[\frac{\tau_{\rm t}}{m_{\rm t}n_{\rm t}} + \frac{\tau_{\rm b}}{m_{\rm b}n_{\rm b}} \right]^{-1}. \tag{5.36}$$

Clearly $\gamma\omega$ is positive for Eq. (5.32), and thus ρ is not singular and a persistent nonzero current (corresponding to $\rho\mathbf{j} = 0$ for $\mathbf{j} \neq 0$) in both layers cannot be maintained without applying external fields. However, if the equality in the above equation is achieved, the double layer carrying persistent currents can exist. This requires that both $\bar{\alpha}$ and β are negative, or $\alpha > 1$ and $\beta < 0$, which will result in a wrong sign of the Coulomb drag resistivity $\rho_{\rm bt}$ and $\rho_{\rm tb}$.

We are particularly interested in the counterflow transport where $\mathbf{j}_{\rm t} = -\mathbf{j}_{\rm b} = \mathbf{j}$. The counterflow field is defined as $\mathbf{E} = \mathbf{E}_{\rm t} - \mathbf{E}_{\rm b} = \rho_{\rm eff}\mathbf{j}$, where the effective resistivity

$$\begin{aligned} \rho_{\rm eff} &= \rho_{\rm t} + \rho_{\rm b} - \rho_{\rm tb} - \rho_{\rm bt} \\ &= \frac{m_{\rm t}}{q_{\rm t}^2 n_{\rm t}\tau_{\rm t}} + \frac{m_{\rm b}}{q_{\rm b}^2 n_{\rm b}\tau_{\rm b}} + \gamma\omega \left(\frac{1}{q_{\rm t}n_{\rm t}} + \frac{1}{q_{\rm b}n_{\rm b}} \right)^2. \end{aligned} \tag{5.37}$$

Thus $\rho_{\rm eff}$ is the sum of the layer resistivity for two layers free from mutual Coulomb interaction (e.g., separated far away), only when $n_{\rm t} = n_{\rm b}$ and the carriers in two layers have opposite charge. The interlayer Coulomb interaction, no matter how strong, will not enter $\rho_{\rm eff}$ at the balanced carrier density, but will however enhance $\rho_{\rm eff}$ at unbalanced or same type of carrier charges of two layers. The classical Drude-like

model tells us that the interlayer Coulomb interaction will not reduce the counterflow resistivity $\rho_{\rm eff}$, and it therefore will not lead to the possibility of the dissipationless conterflow transport. This could be due to the limitations of the classical model, which will be discussed in the following.

The carrier transport in one layer is subjected to the external fields and many scatterings. The carriers are scattered by phonons (acoustic and optical), impurities, defects, surfaces and interfaces, and the other carriers. Those scattering can be either elastic (impurities, defects and acoustic phonons) or inelastic (optical phonons and other carriers). The momentums of the carriers are randomized by the scatterings (no matter elastic or inelastic) so that the momentum relaxation time in the above τ_χ can be defined. It is this momentum relaxation time that is most directly related to the material resistivity. However in the special cases in the double layer structures, the Coulomb scattering from the other layer may not tend to randomize the momentums of carriers. Instead, the momentum transfer will compensate the averaging momentum loss. More critically, the strength of the momentum transfer rate is linearly dependent on the electrical current flow in the other layer. Therefore, a persistent current can be maintained (in a closed circuit) without an external field if the loss of the average momentum is totally compensated by the momentum transfer from the other layer.

Since the total momentum is conserved during the interlayer Coulomb scattering, the momentum transferred from, e.g., the top layer, to the bottom layer is equal to the momentum loss of the top layer. From the perspective of the top layer, this will enhance the resistivity of the top layer, as seen from the term of $\tau_{\rm Dt}$ in $\rho_{\rm t}$, and likewise for the bottom layer. However in quantum theory of solids, the total momentum in the scattering processes is conserved up to reciprocal lattice vectors and the so called Umklapp process is possible. It is then possible to transfer momentum from one layer

to the other without reducing the momentum of the former layer. For example, the sign of the Coulomb drag resistivity can be modulated [184], which cannot be studied using classical models.

Generalizing the above to carriers with zero mass (e.g., electrons/holes in graphene) is a nontrivial task, and it is left for future efforts.

5.4 Boltzmann transport formalism

Within the Boltzmann formalism[4], the transport properties of layer μ can be described by the distribution function $f_\mu(\mathbf{x}, \mathbf{k}_\mu, t)$ for a single band (the band index can actually be absorbed into the index μ), where $\mathbf{k}_\mu$ is the quasi-momentum of the carriers with energy dispersion $e_\mu(\mathbf{k}_\mu)$, $\mathbf{x}$ is the position of electrons, and t is time. The Boltzmann equation (in zero magnetic field) can be formally written as

$$\partial_t f_\mu + \mathbf{v}_\mu(\mathbf{k}_\mu) \cdot \nabla_{\mathbf{x}} f_\mu + \frac{1}{\hbar}\mathbf{F}_\mu(\mathbf{x}) \cdot \nabla_{\mathbf{k}} f_\mu = C_\mu[f_\mu] + \sum_\nu C_{\mu\nu}[f_\mu, f_\nu] \tag{5.38}$$

where C_μ is the single carrier collision integral due to the intralayer scatterings (impurity, defects, phonons etc.), $C_{\mu\nu}$ is the binary collision integral between carriers in layer μ and ν, and the semi-classical dynamics of carriers is

$$\begin{aligned} \partial_t \mathbf{k}_\mu &= \frac{1}{\hbar}\mathbf{F}_\mu = \frac{q_\mu}{\hbar}\mathbf{E}_\mu, \\ \partial_t \mathbf{x} = \mathbf{v}_\mu(\mathbf{k}_\mu) &= \frac{1}{\hbar}\nabla_{\mathbf{k}} e_\mu(\mathbf{k}_\mu) - \frac{1}{\hbar}\mathbf{F}_\mu \times \mathbf{\Omega}_\mu(\mathbf{k}_\mu). \end{aligned} \tag{5.39}$$

Here q_μ is the electrical charge of carriers in layer μ, $\mathbf{E}_\mu$ is the electrical field applied on layer μ, and $\mathbf{\Omega}_\mu(\mathbf{k}_\mu)$ is the Berry curvature of the energy band.

[4]Chapter VII of the textbook by Ziman [185] is referenced when developing this Boltzmann formalism.

The collision integrals can be written as

$$\begin{aligned} C_\mu[f_\mu](\mathbf{k}_\mu) &= \int_{B_\mu} [S_{\rm c}(\mathbf{x},\mathbf{k}'_\mu,\mathbf{k}_\mu) f'_\mu \bar{f}_\mu - S_{\rm c}(\mathbf{x},\mathbf{k}_\mu,\mathbf{k}'_\mu) f_\mu \bar{f}'_\mu] {\rm d}\mathbf{k}'_\mu, \\ C_{\mu\nu}[f_\mu,f_\nu](\mathbf{k}_\mu) &= \int_{B_\mu B_\nu^2} [S_{\rm cc}(\mathbf{x},\mathbf{k}'_\mu,\mathbf{k}_\mu,\mathbf{k}'_\nu,\mathbf{k}_\nu) f'_\mu \bar{f}_\mu f'_\nu \bar{f}_\nu - \\ &\qquad S_{\rm cc}(\mathbf{x},\mathbf{k}_\mu,\mathbf{k}'_\mu,\mathbf{k}_\nu,\mathbf{k}'_\nu) f_\mu \bar{f}'_\mu f_\nu \bar{f}'_\nu] {\rm d}\mathbf{k}'_\mu {\rm d}\mathbf{k}_\nu {\rm d}\mathbf{k}'_\nu, \end{aligned} \tag{5.40}$$

where $S_{\rm c}(\mathbf{x},\mathbf{k}'_\mu,\mathbf{k}_\mu)$ is the transition probability of the single carrier scattering from $\mathbf{k}'_\mu$ to $\mathbf{k}_\mu$, $S_{\rm cc}(\mathbf{x},\mathbf{k}'_\mu,\mathbf{k}_\mu,\mathbf{k}'_\nu,\mathbf{k}_\nu)$ is the transition probability of the two carrier scattering from $\mathbf{k}'_\mu$ to $\mathbf{k}_\mu$ and from $\mathbf{k}'_\nu$ to $\mathbf{k}_\nu$ (the subscript $_{\rm cc}$ refers to carrier-carrier scattering), $\bar{f}_\mu \equiv 1 - f_\mu$, $f'_\mu = f_\mu(\mathbf{x},\mathbf{k}'_\mu,t)$, and B_μ is the reduced (e.g., first) Brillouin zone. The energy and momentum conservation laws and the quantum mechanical nature of the scatterings are implicitly encoded in the mathematical forms of $S_{\rm c}$ and $S_{\rm cc}$. Note that in these integrals $\mathbf{k}_\nu$ and $\mathbf{k}'_\nu$ have to be replaced by, e.g., $\mathbf{k}''_\mu$ and $\mathbf{k}'''_\mu$ respectively, when $\nu = \mu$.

After linearizing the Boltzmann equation (see Appendix G.1), the left hand side of Eq. (5.38) is

$$X_\mu \equiv \frac{1}{\hbar}\mathbf{F}_\mu \cdot \nabla_{\mathbf{k}} n_\mu = -\frac{1}{k_{\rm B}T} n_\mu \bar{n}_\mu \mathbf{F}_\mu \cdot \mathbf{v}_\mu \tag{5.41}$$

up to the first order in ϕ_μ. The linearized Boltzmann equations now become

$$X_\mu = \sum_\nu P_{\mu\nu}\phi_\nu \tag{5.42}$$

where the P operators are defined as

$$\begin{aligned} P_{\mu\nu}\phi_\nu \equiv \delta_{\mu\nu} &\int_{B_\mu} W(\mu'\mu)[\phi_\mu - \phi'_\mu]{\rm d}\mu' \\ &+ \int_{B_\mu B_\nu^2} Q_{\rm cc}(\mu'\mu\nu'\nu)[\phi_\nu - \phi'_\nu]{\rm d}\mu'{\rm d}\nu{\rm d}\nu', \end{aligned} \tag{5.43}$$

with

$$W(\mu'\mu) \equiv Q_{\rm c}(\mu'\mu) + \sum_\nu \int_{B_\nu^2} Q_{\rm cc}(\mu'\mu\nu'\nu){\rm d}\nu{\rm d}\nu'. \tag{5.44}$$

We obviously have $W(\mu'\mu) = W(\mu\mu')$. By using the symmetry properties of W and $Q_{\rm cc}$, we have

$$\langle \psi_\mu, P_{\mu\nu}\phi_\nu \rangle = \langle \phi_\nu, P_{\nu\mu}\psi_\mu \rangle \,, \tag{5.45}$$

for any given ψ_μ, where the inner product is defined as

$$\langle \psi_\mu, \varphi_\mu \rangle \equiv \int_{B_\mu} \psi_\mu(\mathbf{k}_\mu)\varphi_\mu(\mathbf{k}_\mu)\mathrm{d}\mu, \tag{5.46}$$

since

$$\begin{aligned} \langle \psi_\mu, P_{\mu\nu}\phi_\nu \rangle = {} & \frac{1}{2}\delta_{\mu\nu}\int_{B_\mu^2} W(\mu'\mu)[\psi_\mu - \psi'_\mu][\phi_\mu - \phi'_\mu]\mathrm{d}\mu\mathrm{d}\mu' \\ & + \frac{1}{2}\int_{B_\mu^2 B_\nu^2} Q_{\rm cc}(\mu'\mu\nu'\nu)[\psi_\mu - \psi'_\mu][\phi_\nu - \phi'_\nu]\mathrm{d}\mu\mathrm{d}\mu'\mathrm{d}\nu\mathrm{d}\nu'. \end{aligned} \tag{5.47}$$

For any two given function vectors $\mathbf{\Phi}$ and $\mathbf{\Psi}$ with components ϕ_μ and ψ_μ respectively, we have

$$\langle \mathbf{\Psi}, P\mathbf{\Phi} \rangle \equiv \sum_{\mu\nu} \langle \psi_\mu, P_{\mu\nu}\phi_\nu \rangle = \langle \mathbf{\Phi}, P\mathbf{\Psi} \rangle \,, \tag{5.48}$$

where the matrix operator P is defined in such a way that the μ component of $P\mathbf{\Phi}$ is $(P\mathbf{\Phi})_\mu = \sum_\nu P_{\mu\nu}\phi_\nu$. On the other hand,

$$\begin{aligned} \langle \mathbf{\Phi}, P\mathbf{\Phi} \rangle = {} & \frac{1}{2}\sum_\mu \int_{B_\mu^2} W(\mu'\mu)[\phi_\mu - \phi'_\mu]^2\mathrm{d}\mu\mathrm{d}\mu' \\ & + \frac{1}{2}\sum_{\mu\nu}\int_{B_\mu^2 B_\nu^2} Q_{\rm cc}(\mu'\mu\nu'\nu)[\phi_\mu - \phi'_\mu][\phi_\nu - \phi'_\nu]\mathrm{d}\mu\mathrm{d}\mu'\mathrm{d}\nu\mathrm{d}\nu' \\ = {} & \frac{1}{2}\sum_\mu \int_{B_\mu^2} Q_{\rm c}(\mu'\mu)[\phi_\mu - \phi'_\mu]^2\mathrm{d}\mu\mathrm{d}\mu' \\ & + \frac{1}{4}\sum_{\mu\nu}\int_{B_\mu^2 B_\nu^2} Q_{\rm cc}(\mu'\mu\nu'\nu)[\phi_\mu - \phi'_\mu + \phi_\nu - \phi'_\nu]^2\mathrm{d}\mu\mathrm{d}\mu'\mathrm{d}\nu\mathrm{d}\nu'. \end{aligned} \tag{5.49}$$

Therefore $\langle \mathbf{\Phi}, P\mathbf{\Phi} \rangle \geq 0$ and thus P is positive semi-definite and self-adjoint. We can write down the formal solution of the Boltzmann equation as

$$\mathbf{\Phi} = L\mathbf{X} \tag{5.50}$$

where the μ component of $\mathbf{X}$ is X_μ and L is the inverse operator of P, i.e., $PL = LP = \mathcal{I}$, with $\mathcal{I}$ being the identity operator.

Physically, we can reasonably assume that a nonzero input $\mathbf{X}$ will generate a unique nonzero output $\mathbf{\Phi}$. Therefore, L physically exists on the space spanned by all of the components of all $\nabla_{\mathbf{k}} n_\mu$.

5.4.1 Onsager relation of Coulomb drag conductivity

Since P is linear and self-adjoint, so is L. Then the above solution can be expanded as

$$\phi_\mu = \sum_\nu L_{\mu\nu} X_\nu = \frac{1}{\hbar} \sum_\nu \mathbf{F}_\nu \cdot L_{\mu\nu} \nabla_{\mathbf{k}} n_\nu. \tag{5.51}$$

The electrical current is

$$\begin{aligned}
\mathbf{j}_\mu &= q_\mu \int_{B_\mu} \mathbf{v}_\mu(\mathbf{k}_\mu) f_\mu \mathrm{d}\mu \\
&= \frac{q_\mu}{\hbar} \left[\int_{B_\mu} n_\mu \mathbf{\Omega}_\mu \mathrm{d}\mu \right] \times \mathbf{F}_\mu - \frac{q_\mu}{\hbar} \int_{B_\mu} (\nabla_{\mathbf{k}} e_\mu(\mathbf{k}_\mu)) n_\mu \bar{n}_\mu \phi_\mu \mathrm{d}\mu \\
&= \frac{q_\mu}{\hbar} \left[\int_{B_\mu} n_\mu \mathbf{\Omega}_\mu \mathrm{d}\mu \right] \times \mathbf{F}_\mu + \frac{k_\mathrm{B} T q_\mu}{\hbar^2} \int_{B_\mu} \nabla_{\mathbf{k}} n_\mu \sum_\nu \mathbf{F}_\nu \cdot L_{\mu\nu} \nabla_{\mathbf{k}} n_\nu \mathrm{d}\mu \\
&= \frac{q_\mu}{\hbar} \left[\int_{B_\mu} n_\mu \mathbf{\Omega}_\mu \mathrm{d}\mu \right]_\times \mathbf{F}_\mu + \frac{k_\mathrm{B} T q_\mu}{\hbar^2} \sum_\nu \left[\int_{B_\mu} (\nabla_{\mathbf{k}} n_\mu) \otimes L_{\mu\nu} \nabla_{\mathbf{k}} n_\nu \mathrm{d}\mu \right] \mathbf{F}_\nu,
\end{aligned} \tag{5.52}$$

where we have used the notations

$$\begin{aligned}
[\mathbf{a}]_\times \mathbf{b} &\equiv \mathbf{a} \times \mathbf{b}, \\
(\mathbf{a} \otimes \mathbf{b})_{mn} &\equiv a_m b_n,
\end{aligned} \tag{5.53}$$

in which $\mathbf{a}$ and $\mathbf{b}$ are vectors in the Cartesian coordinate and their components are labeled by a_m and b_m respectively. The conductivity matrix can then be written as

$$\sigma_{\mu\nu} = \delta_{\mu\nu} \frac{q_\mu^2}{\hbar} \left[\int_{B_\mu} n_\mu \mathbf{\Omega}_\mu \mathrm{d}\mu \right]_\times + \frac{k_\mathrm{B} T q_\mu q_\nu}{\hbar^2} \int_{B_\mu} (\nabla_{\mathbf{k}} n_\mu) \otimes L_{\mu\nu} \nabla_{\mathbf{k}} n_\nu \mathrm{d}\mu. \tag{5.54}$$

Since the Berry curvature term will not affect the interlayer transport, we will ignore it. The conductivity matrix is then

$$\sigma_{\mu\nu} = \frac{k_{\mathrm{B}} T q_\mu q_\nu}{\hbar^2} \int_{B_\mu} (\nabla_{\mathbf{k}} n_\mu) \otimes (L_{\mu\nu} \nabla_{\mathbf{k}} n_\nu) \mathrm{d}\mu. \tag{5.55}$$

We then have obtained the Onsager relation for the conductivity matrices:

$$\sigma_{\mu\nu} = [\sigma_{\nu\mu}]^T, \tag{5.56}$$

because L is self-adjoint.

5.4.2 Entropy production

The Joule heating generation in layer μ is

$$\begin{aligned} \mathbf{j}_\mu \cdot \mathbf{E}_\mu &= \sum_\nu \frac{k_{\mathrm{B}} T}{\hbar^2} \int_{B_\mu} (\mathbf{F}_\mu \cdot \nabla_{\mathbf{k}} n_\mu)(L_{\mu\nu} \mathbf{F}_\nu \cdot \nabla_{\mathbf{k}} n_\nu) \mathrm{d}\mu \\ &= k_{\mathrm{B}} T \int_{B_\mu} X_\mu \phi_\mu \mathrm{d}\mu \\ &= k_{\mathrm{B}} T \left\langle \phi_\mu, X_\mu \right\rangle, \end{aligned} \tag{5.57}$$

and the total Joule heating generation is

$$\sum_\mu \mathbf{j}_\mu \cdot \mathbf{E}_\mu = k_{\mathrm{B}} T \left\langle \mathbf{\Phi}, P\mathbf{\Phi} \right\rangle, \tag{5.58}$$

where $k_{\mathrm{B}} \left\langle \mathbf{\Phi}, P\mathbf{\Phi} \right\rangle$ is the entropy production rate due to carrier transport, and is exactly balanced by the Joule heating dissipation (see Appendix G.2).

We now want to know how the existence of interlayer interactions affect the entropy production rate. To proceed, we start with the following settings. Let Q be the linear and self-adjoint operator that has the same form as P excluding all the terms of $Q_{\mathrm{cc}}(\mu'\mu\nu'\nu)$ for $\mu \neq \nu$, i.e., Q and P describe the same system where in Q

the interlayer interactions are turned off. We assume the equilibrium distributions are same for the Q and P systems. We now applied the same electrical fields to both systems, and the responding deviations of the equilibrium distributions are $\mathbf{\Psi}$ and $\mathbf{\Phi}$ respectively. They are related by

$$Q\mathbf{\Psi} = \mathbf{X} = P\mathbf{\Phi}. \tag{5.59}$$

We want to know if the entropy production rate $k_{\mathrm{B}} \langle \mathbf{\Phi}, P\mathbf{\Phi} \rangle$ is smaller than $k_{\mathrm{B}} \langle \mathbf{\Psi}, Q\mathbf{\Psi} \rangle$. We have

$$\begin{aligned} &\langle \mathbf{\Psi}, Q\mathbf{\Psi} \rangle - \langle \mathbf{\Phi}, P\mathbf{\Phi} \rangle = \langle \mathbf{\Psi}, (P-Q)\mathbf{\Phi} \rangle \\ &= \frac{1}{4} \sum_{\mu \neq \nu} \int_{B_\mu^2 B_\nu^2} Q_{\mathrm{cc}}(\mu' \mu \nu' \nu)[\psi_\mu - \psi'_\mu + \psi_\nu - \psi'_\nu][\phi_\mu - \phi'_\mu + \phi_\nu - \phi'_\nu] \mathrm{d}\mu \mathrm{d}\mu' \mathrm{d}\nu \mathrm{d}\nu'. \end{aligned} \tag{5.60}$$

Apparently $P - Q$ is self-adjoint and positive semi-definite. On the other hand, if we assume the inverse of Q and P are existed and they are denoted by Q^{-1} and P^{-1} respectively, we have

$$\begin{aligned} \langle \mathbf{\Psi}, Q\mathbf{\Psi} \rangle - \langle \mathbf{\Phi}, P\mathbf{\Phi} \rangle &= \left\langle Q^{-1}\mathbf{X}, (P-Q)P^{-1}\mathbf{X} \right\rangle \\ &= \left\langle Q^{-1}(P-Q)P^{-1}\mathbf{X}, \mathbf{X} \right\rangle \geq 0, \end{aligned} \tag{5.61}$$

where the positive semi-definiteness of $Q^{-1}(P-Q)P^{-1}$ is used. Thus the existence of the interlayer Coulomb interactions will not increase the entropy production rate.

So far the above analysis can also be applied to a system containing carriers with species labeled by μ where the carrier number of each species is conserved. The inter-species Coulomb interactions will thus reduce the entropy production rate. In other words, by turning on the inter-species Coulomb interactions, the system resistivity will increase, in agreement with the intuitive picture that more scatterings with lead to a more resistive state.[5] However, the inter-species interaction can not be easily

[5]The carrier-carrier scattering may not contribute to the resistivity if the scattering is not Umklapp and the carriers have quadratic dispersion with same effective mass.

tuned. In contrast, the interlayer Coulomb interactions can be tuned by the interlayer distance.

5.4.3 Boltzmann transport in presence of excitons

When the interlayer excitons are formed for carriers in layer μ and ν, its distribution function is written as $f_{\mu\nu}(x, \mathbf{k}_{\mu\nu}, t)$, where $\mathbf{k}_{\mu\nu}$ is the quasi-momentum of the excitons with energy dispersion $e_{\mu\nu}(\mathbf{k}_{\mu\nu})$. Now the coupled Boltzmann equations are

$$\begin{aligned}
&\partial_t f_\mu + \mathbf{v}_\mu(\mathbf{k}_\mu)\cdot\nabla_{\mathbf{x}} f_\mu + \frac{\mathbf{F}_\mu}{\hbar}\cdot\nabla_{\mathbf{k}} f_\mu = \\
&\qquad C_\mu[f_\mu] + \sum_\nu \left(C_{\mu\nu}[f_\mu, f_\nu] + D_{\mu\nu}[f_\mu, f_\nu, f_{\mu\nu}]\right) + \sum_{\alpha\beta} D_{\alpha\beta,\mu}[f_{\alpha\beta}, f_\mu], \\
&\partial_t f_{\mu\nu} + \mathbf{v}_{\mu\nu}(\mathbf{k}_{\mu\nu})\cdot\nabla_{\mathbf{x}} f_{\mu\nu} + \frac{\mathbf{F}_\mu + \mathbf{F}_\nu}{\hbar}\cdot\nabla_{\mathbf{k}} f_{\mu\nu} = \\
&\qquad C^{\mathrm{x}}_{\mu\nu}[f_{\mu\nu}] + D^{\mathrm{x}}_{\mu\nu}[f_\mu, f_\nu, f_{\mu\nu}] + \sum_{\alpha\beta} C^{\mathrm{x}}_{\mu\nu,\alpha\beta}[f_{\mu\nu}, f_{\alpha\beta}] + \sum_\alpha D^{\mathrm{x}}_{\mu\nu,\alpha}[f_{\mu\nu}, f_\alpha],
\end{aligned} \tag{5.62}$$

where the collision integrals C_μ and $C_{\mu\nu}$ are defined as in Eq. (5.40), and the other integrals are defined in Appendix G.4.

The linearized Boltzmann equations can now be written as

$$X_{\mu\nu} \equiv \frac{1}{\hbar}\mathbf{F}_{\mu\nu}\cdot\nabla_{\mathbf{k}} n_{\mu\nu} = \sum_{\alpha\beta} P_{\mu\nu,\alpha\beta}\phi_{\alpha\beta}, \tag{5.63}$$

where

$$\mathbf{F}_{\mu\nu} \equiv \frac{\mathbf{F}_\mu + \mathbf{F}_\nu}{1+\delta_{\mu\nu}} = \frac{q_\mu \mathbf{E}_\mu + q_\nu \mathbf{E}_\nu}{1+\delta_{\mu\nu}},\ n_{\mu\mu} = n_\mu,\ \phi_{\mu\mu} = \phi_\mu. \tag{5.64}$$

The linear operators $P_{\mu\nu,\alpha\beta}$ are shown in Appendix G.4.1.

The symmetry property of the P operators is spoiled by the processes of the creation/annihilation of excitons, because

$$\begin{aligned}\langle\psi_\mu, P_{\mu\mu,\alpha\beta}\phi_{\alpha\beta}\rangle - \langle\phi_{\alpha\beta}, P_{\alpha\beta,\mu\mu}\psi_\mu\rangle = \delta_{\mu\beta}\int_{B_\mu B_\alpha B_{\alpha\beta}} Q_{\rm xca}([\alpha\beta]\alpha\mu)\psi_\mu\phi_{\alpha\beta}{\rm d}\alpha{\rm d}\mu{\rm d}[\alpha\beta]\\ \neq 0.\end{aligned} \tag{5.65}$$

Therefore, the P operator is not self-adjoint. If we let $\mathbf{\Phi}$ be the vector with components $\phi_{\mu\nu}$ and $\mathbf{X}$ be the vector with components $X_{\mu\nu}$, we can define

$$\begin{aligned}\langle\mathbf{\Phi}, P\mathbf{\Phi}\rangle \equiv &\sum_{\mu\nu}\langle\phi_{\mu\nu}, X_{\mu\nu}\rangle = \sum_{\mu\nu,\alpha\beta}\langle\phi_{\mu\nu}, P_{\mu\nu,\alpha\beta}\phi_{\alpha\beta}\rangle\\ =&R_{\rm c} + R_{\rm x} + R_{\rm xc} + R_{\rm xca},\end{aligned} \tag{5.66}$$

where the contribution from carriers is

$$\begin{aligned}R_{\rm c} =&\frac{1}{2}\sum_\mu\int_{B_\mu^2} Q_{\rm c}(\mu'\mu)[\phi_\mu - \phi'_\mu]^2{\rm d}\mu{\rm d}\mu' +\\ &\frac{1}{4}\sum_{\mu\nu}\int_{B_\mu^2 B_\nu^2} Q_{\rm cc}(\mu'\mu\nu'\nu)[\phi_\mu - \phi'_\mu + \phi_\nu - \phi'_\nu]^2{\rm d}\mu{\rm d}\mu'{\rm d}\nu{\rm d}\nu',\end{aligned} \tag{5.67}$$

the contribution from excitons is

$$\begin{aligned}R_{\rm x} =&\frac{1}{2}\sum_{\mu\neq\nu}\int_{B_{\mu\nu}^2} Q_{\rm x}([\mu\nu]'[\mu\nu])[\phi_{\mu\nu} - \phi'_{\mu\nu}]^2{\rm d}[\mu\nu]{\rm d}[\mu\nu]' +\\ &\frac{1}{4}\sum_{\substack{\mu\neq\nu\\ \alpha\neq\beta}}\int_{\substack{B_{\mu\nu}^2\\ B_{\alpha\beta}^2}} Q_{\rm xx}([\mu\nu]'[\mu\nu][\alpha\beta]'[\alpha\beta])\begin{bmatrix}\phi_{\mu\nu}-\phi'_{\mu\nu}+\\ \phi_{\alpha\beta}-\phi'_{\alpha\beta}\end{bmatrix}^2{\rm d}[\mu\nu]{\rm d}[\mu\nu]'{\rm d}[\alpha\beta]{\rm d}[\alpha\beta]',\end{aligned} \tag{5.68}$$

the contribution from exciton-carrier interaction is

$$R_{\rm xc} = \frac{1}{4}\sum_{\mu\neq\nu,\alpha}\int_{B_{\mu\nu}^2 B_\alpha^2} Q_{\rm xc}([\mu\nu]'[\mu\nu]\alpha'\alpha)\begin{bmatrix}\phi_{\mu\nu}-\phi'_{\mu\nu}+\\ \phi_\alpha-\phi'_\alpha\end{bmatrix}^2{\rm d}[\mu\nu]{\rm d}[\mu\nu]'{\rm d}\alpha{\rm d}\alpha', \tag{5.69}$$

and the contribution from the exciton creation/annihilation is

$$R_{\rm xca} = \sum_{\mu\neq\nu}\int_{B_{\mu\nu}B_\mu B_\nu} Q_{\rm xca}([\mu\nu]\mu\nu)\left[\phi_\mu - \phi_{\mu\nu}\right]\left[\phi_\mu + \phi_\nu - \phi_{\mu\nu}\right]{\rm d}[\mu\nu]{\rm d}\mu{\rm d}\nu. \tag{5.70}$$

Due to the presence of $R_{\rm xca}$ which might be negative, the operator P is no longer guaranteed to be positive semi-definite.

We again assume that the inverse of P exists, i.e., $P^{-1} = L$. The formal solution of the Boltzmann equations can be written as

$$\phi_{\mu\nu} = \sum_{\alpha\beta} L_{\mu\nu,\alpha\beta} X_{\alpha\beta} = \frac{1}{\hbar} \sum_{\alpha\beta} \mathbf{F}_{\alpha\beta} \cdot L_{\mu\nu,\alpha\beta} \nabla_{\mathbf{k}} n_{\alpha\beta}. \tag{5.71}$$

The electrical current is

$$\begin{aligned}
\mathbf{j}_\mu &= q_\mu \sum_\nu \int_{B_{\mu\nu}} \mathbf{v}_{\mu\nu}(\mathbf{k}_{\mu\nu}) f_{\mu\nu} \mathrm{d}[\mu\nu] = -q_\mu \sum_\nu \int_{B_{\mu\nu}} \mathbf{v}_{\mu\nu}(\mathbf{k}_{\mu\nu}) n_{\mu\nu} \bar{n}_{\mu\nu} \phi_{\mu\nu} \mathrm{d}[\mu\nu] \\
&= -\frac{q_\mu}{\hbar} \sum_{\nu\alpha\beta} \int_{B_{\mu\nu}} \mathbf{v}_{\mu\nu}(\mathbf{k}_{\mu\nu}) n_{\mu\nu} \bar{n}_{\mu\nu} \mathbf{F}_{\alpha\beta} \cdot L_{\mu\nu,\alpha\beta} \nabla_{\mathbf{k}} n_{\alpha\beta} \mathrm{d}[\mu\nu] \\
&= \frac{k_{\rm B} T q_\mu}{\hbar^2} \sum_{\nu\alpha\beta} \int_{B_{\mu\nu}} \nabla_{\mathbf{k}} n_{\mu\nu} \mathbf{F}_{\alpha\beta} \cdot L_{\mu\nu,\alpha\beta} \nabla_{\mathbf{k}} n_{\alpha\beta} \mathrm{d}[\mu\nu] \\
&= \frac{k_{\rm B} T q_\mu}{\hbar^2} \sum_{\nu\alpha\beta} \frac{1}{1+\delta_{\alpha\beta}} \left[\int_{B_{\mu\nu}} \nabla_{\mathbf{k}} n_{\mu\nu} \otimes L_{\mu\nu,\alpha\beta} \nabla_{\mathbf{k}} n_{\alpha\beta} \mathrm{d}[\mu\nu] \right] (\mathbf{F}_\alpha + \mathbf{F}_\beta).
\end{aligned} \tag{5.72}$$

The conductivity matrix can then be formally written as

$$\sigma_{\mu\nu} = \frac{2 k_{\rm B} T q_\mu q_\nu}{\hbar^2} \sum_{\alpha\beta} \frac{1}{1+\delta_{\beta\nu}} \int_{B_{\mu\alpha}} \nabla_{\mathbf{k}} n_{\mu\alpha} \otimes L_{\mu\alpha,\nu\beta} \nabla_{\mathbf{k}} n_{\nu\beta} \mathrm{d}[\mu\alpha]. \tag{5.73}$$

If the exciton annihilation/creation processes are ignored, i.e., $Q_{\rm xca} = 0$, P is self-adjoint and positive semi-definite and thus L is also self-adjoint. The Onsager relations, as in Eq. (5.56), will hold if

$$\sum_\beta \int_{B_\mu} \nabla_{\mathbf{k}} n_\mu \otimes L_{\mu\mu,\nu\beta} \nabla_{\mathbf{k}} n_{\nu\beta} \mathrm{d}\mu = \sum_\alpha \int_{B_{\mu\alpha}} \nabla_{\mathbf{k}} n_{\mu\alpha} \otimes L_{\mu\alpha,\nu\nu} \nabla_{\mathbf{k}} n_\nu \mathrm{d}[\mu\alpha] \tag{5.74}$$

is satisfied. For a two layer system, i.e., $\mu, \nu \in \{\mathrm{t}, \mathrm{b}\}$, the above is equivalent to

$$\int_{B_{\rm t}} \nabla_{\mathbf{k}} n_{\rm t} \otimes L_{\rm tt,tb} \nabla_{\mathbf{k}} n_{\rm tb} \mathrm{dt} = \int_{B_{\rm b}} \nabla_{\mathbf{k}} n_{\rm b} \otimes L_{\rm bb,tb} \nabla_{\mathbf{k}} n_{\rm tb} \mathrm{db}. \tag{5.75}$$

Since the chemical potential $\zeta_{\mathrm{t/b}}$ can be varied arbitrarily, and the operator L is a functional of $n_{\mathrm{t,b,tb}}$, rather than the gradient of them, the above equality cannot be always guaranteed. Therefore, the existence of exciton-carrier scattering alone can break the Onsager relation. However, if the scattering between excitons and carriers can be ignored, i.e., the transport channels of carriers are independent of the transport channels of excitons, we have $L_{\mu\mu,\alpha\beta} = L_{\alpha\beta,\mu\mu} = 0$ because $P_{\mu\mu,\alpha\beta} = P_{\alpha\beta,\mu\mu} = 0$, and therefore the Onsager relation holds.

The total Joule heating generation is

$$\begin{aligned}
\sum_{\mu} \mathbf{j}_{\mu} \cdot \mathbf{E}_{\mu} &= k_{\mathrm{B}} T \sum_{\mu\nu} \int_{B_{\mu\nu}} \phi_{\mu\nu} \frac{\mathbf{F}_{\mu} \cdot \nabla_{\mathbf{k}} n_{\mu\nu}}{\hbar} \mathrm{d}[\mu\nu] \\
&= \frac{1}{2} k_{\mathrm{B}} T \sum_{\mu\nu} \int_{B_{\mu\nu}} \phi_{\mu\nu} \frac{(\mathbf{F}_{\mu} + \mathbf{F}_{\nu}) \cdot \nabla_{\mathbf{k}} n_{\mu\nu}}{\hbar} \mathrm{d}[\mu\nu] \\
&= \frac{1}{2} k_{\mathrm{B}} T \sum_{\mu\nu} \langle \phi_{\mu\nu}, X_{\mu\nu} \rangle (1 + \delta_{\mu\nu}) \\
&= \frac{1}{2} k_{\mathrm{B}} T \left\langle \mathbf{\Phi}, P\mathbf{\Phi} \right\rangle + \frac{1}{2} k_{\mathrm{B}} T \sum_{\mu} \langle \phi_{\mu}, X_{\mu} \rangle \\
&= k_{\mathrm{B}} T \left\langle \mathbf{\Phi}, P\mathbf{\Phi} \right\rangle - \frac{1}{2} k_{\mathrm{B}} T \sum_{\mu \neq \nu} \langle \phi_{\mu\nu}, X_{\mu\nu} \rangle,
\end{aligned} \tag{5.76}$$

which is not equal to $k_{\mathrm{B}} T \left\langle \mathbf{\Phi}, P\mathbf{\Phi} \right\rangle$ in general. However if only intralayer excitons are considered which is equivalent to $\mathbf{F}_{\mu\nu} = 0$ in our formulations, the total Joule heat generation is equal to $k_{\mathrm{B}} T \left\langle \mathbf{\Phi}, P\mathbf{\Phi} \right\rangle$.

Consider two extreme cases. If there are no excitons, we have $\sum_{\mu} \mathbf{j}_{\mu} \cdot \mathbf{E}_{\mu} = k_{\mathrm{B}} T \left\langle \mathbf{\Phi}, P\mathbf{\Phi} \right\rangle$, same as the previous sections. If only excitons exist without free carriers, we have $\sum_{\mu} \mathbf{j}_{\mu} \cdot \mathbf{E}_{\mu} = \frac{1}{2} k_{\mathrm{B}} T \left\langle \mathbf{\Phi}, P\mathbf{\Phi} \right\rangle$. The factor $\frac{1}{2}$ is due to the fact that exciton distribution function $\phi_{\mu\nu}$. Now the conductivity matrix is

$$\sigma_{\mu\nu} = \frac{2 k_{\mathrm{B}} T q_{\mu} q_{\nu}}{\hbar^2} \sum_{\alpha \neq \mu,\ \beta \neq \nu} \int_{B_{\mu\alpha}} \nabla_{\mathbf{k}} n_{\mu\alpha} \otimes L_{\mu\alpha,\nu\beta} \nabla_{\mathbf{k}} n_{\nu\beta} \mathrm{d}[\mu\alpha] \tag{5.77}$$

If there are only two layers, i.e., $\mu, \nu \in \{\mathrm{t}, \mathrm{b}\}$, we have

$$\sigma_{\mathrm{tt}} = \sigma_{\mathrm{bb}} = -\sigma_{\mathrm{tb}} = -\sigma_{\mathrm{bt}} = \sigma_0 = \frac{2k_{\mathrm{B}}Tq^2}{\hbar^2} \int_{B_{\mathrm{tb}}} \nabla_{\mathbf{k}} n_{\mathrm{tb}} \otimes L_{\mathrm{tb,tb}} \nabla_{\mathbf{k}} n_{\mathrm{tb}} \mathrm{d}[\mathrm{tb}], \tag{5.78}$$

where $q_{\mathrm{t}} = -q_{\mathrm{b}} = q$. For the isotropic case, σ_0 is a scalar. The total Joule heating now is

$$\mathbf{j}_{\mathrm{t}} \cdot \mathbf{E}_{\mathrm{t}} + \mathbf{j}_{\mathrm{b}} \cdot \mathbf{E}_{\mathrm{b}} = \sigma_0 (\mathbf{E}_{\mathrm{t}} - \mathbf{E}_{\mathrm{b}})^2, \tag{5.79}$$

which goes to zero if $\mathbf{E}_{\mathrm{t}} = \mathbf{E}_{\mathrm{b}}$. Then we would have $\mathbf{F}_{\mathrm{tb}} = 0$, and thus the exciton is not driven out of equilibrium and the current $\mathbf{j}_{\mathrm{t}}$ and $\mathbf{j}_{\mathrm{b}}$ are zero. The conductivity matrix is

$$\sigma = \sigma_0 \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}. \tag{5.80}$$

Apparently, σ is singular and it cannot be inverted to obtain the resistivity matrix. In fact, this structure of matrix is a consequence of exciton conduction where excitons are stable composite particles. The conductivity σ_0 should be explained as the conductivity of excitons, rather than individual layer or Coulomb drag conductivity. The exciton resistivity is then $\rho_0 = \sigma_0^{-1}$. Since the electrical currents in two layers are not independent, a resistivity matrix cannot be defined.

For the two layer system, we assume now the carrier conduction and exciton conduction are independent. The conductivity matrix can be written as

$$\sigma = \begin{bmatrix} \sigma_0 + \sigma_{\mathrm{t}} & -(\sigma_0 + \sigma_{\mathrm{d}}) \\ -(\sigma_0 + \sigma_{\mathrm{d}}) & \sigma_0 + \sigma_{\mathrm{b}} \end{bmatrix} \tag{5.81}$$

where

$$\begin{aligned} \sigma_{\mathrm{t}} &= \frac{k_{\mathrm{B}}Tq^2}{\hbar^2} \int_{B_{\mathrm{tt}}} \nabla_{\mathbf{k}} n_{\mathrm{tt}} \otimes L_{\mathrm{tt,tt}} \nabla_{\mathbf{k}} n_{\mathrm{tt}} \mathrm{d}[\mathrm{tt}], \\ \sigma_{\mathrm{b}} &= \frac{k_{\mathrm{B}}Tq^2}{\hbar^2} \int_{B_{\mathrm{bb}}} \nabla_{\mathbf{k}} n_{\mathrm{bb}} \otimes L_{\mathrm{bb,bb}} \nabla_{\mathbf{k}} n_{\mathrm{bb}} \mathrm{d}[\mathrm{bb}], \\ \sigma_{\mathrm{d}} &= \frac{k_{\mathrm{B}}Tq^2}{\hbar^2} \int_{B_{\mathrm{tt}}} \nabla_{\mathbf{k}} n_{\mathrm{tt}} \otimes L_{\mathrm{tt,bb}} \nabla_{\mathbf{k}} n_{\mathrm{bb}} \mathrm{d}[\mathrm{tt}]. \end{aligned} \tag{5.82}$$

We can again assume isotropic transport, i.e., σ_0, σ_t, σ_b and σ_d are all scalars. The total Joule heat generation in counterflow transport is now

$$\mathbf{j}_\mathrm{t}\cdot\mathbf{E}_\mathrm{t}+\mathbf{j}_\mathrm{b}\cdot\mathbf{E}_\mathrm{b}=\sigma_0\mathbf{E}^2+\frac{\sigma_\mathrm{t}\sigma_\mathrm{b}-\sigma_\mathrm{d}^2}{\sigma_\mathrm{t}+\sigma_\mathrm{b}-2\sigma_\mathrm{d}}\mathbf{E}^2, \tag{5.83}$$

where we have required $\mathbf{j}_\mathrm{t}+\mathbf{j}_\mathrm{b}=0$ and $\mathbf{E}=\mathbf{E}_\mathrm{t}-\mathbf{E}_\mathrm{b}$. Apparently the heat generation in the carrier channel is independent of exciton channel, as expected. The conductivity matrix σ can be inverted to obtain the resistivity matrix

$$\rho=\frac{1}{\sigma_\Delta}\begin{bmatrix}\sigma_0+\sigma_\mathrm{b} & \sigma_0+\sigma_\mathrm{d}\\ \sigma_0+\sigma_\mathrm{d} & \sigma_0+\sigma_\mathrm{t}\end{bmatrix} \tag{5.84}$$

where $\sigma_\Delta=\sigma_\mathrm{t}\sigma_\mathrm{b}-\sigma_\mathrm{d}^2+\sigma_0(\sigma_\mathrm{t}+\sigma_\mathrm{b}-2\sigma_\mathrm{d})$. We can identify

$$\rho_\mathrm{t}=\frac{\sigma_0+\sigma_\mathrm{b}}{\sigma_\Delta},\quad \rho_\mathrm{b}=\frac{\sigma_0+\sigma_\mathrm{t}}{\sigma_\Delta},\quad \rho_\mathrm{d}=\frac{\sigma_0+\sigma_\mathrm{d}}{\sigma_\Delta}. \tag{5.85}$$

Even though σ_d could be much less than $\sigma_{\mathrm{t/b}}$, we could still have ρ_d much close to $\rho_{\mathrm{t/b}}$ if σ_0 is dominantly large (e.g., when the exciton density is much larger than the carrier density). To see this, we can drop the first two terms in σ_Δ and also drop σ_d. Now the resistivity matrix is

$$\rho=\frac{1}{\sigma_\mathrm{t}+\sigma_\mathrm{b}}\begin{bmatrix}1+\dfrac{\sigma_\mathrm{b}}{\sigma_0} & 1\\ 1 & 1+\dfrac{\sigma_\mathrm{t}}{\sigma_0}\end{bmatrix}. \tag{5.86}$$

The round bracket in Eq. (5.8) now is $\rho_\mathrm{t}+\rho_\mathrm{b}-\rho_\mathrm{tb}-\rho_\mathrm{bt}=\sigma_0^{-1}=\rho_0$. Thus, high ZT actually requires large exciton conductivity, since apparently exciton conduction is dominant in this case.

5.5 General multilayer system

For closely stacked but electrically isolated N layers of uniform two-dimensional conducting systems in Fig. 5.6, the current flow in one layer will produce electrical

field on the layer due to Ohm's law and Coulomb drag electrical field on the other layers due to Coulomb force coupling between layers. We assume that the current flow in each layer is uniform and in the same direction. The currents and electrical fields can then be characterized by scalars J_i and E_i respectively for $i = 1, \cdots, N$, where the layers are indexed by integers i with 1 being the top-most layer and N being the bottom-most layer. Let $\boldsymbol{J} = (J_1, J_2, \cdots, J_N)$ and $\boldsymbol{E} = (E_1, E_2, \cdots, E_N)$. They are connected by the $N \times N$ resistivity matrix $\rho = [\rho_{i,j}]$ where $\rho_{i,j}$ $(i \neq j)$ is the drag resistivity when the current flows in layer j while the electrical field is measured in layer i and the diagonal entries $\rho_{i,i} = \rho_i$ are the intralayer resistivities, i.e., $\boldsymbol{E} = \rho \boldsymbol{J}^T$ where the superscript T represents the transpose operation. Here we consider the possibility that the Coulomb drag resistivity violates the Onsager's relation [172], i.e., $\rho_{i,j}$ and $\rho_{j,i}$ can be different. We limit our discussions in the linear response regime. Usually only the Coulomb drag between two adjacent layers are considered due to the strong screening effect. The resistivity matrix then becomes tridiagonal.

The total power consumption is $P = \boldsymbol{J}\rho\boldsymbol{J}^T = \boldsymbol{J}A\boldsymbol{J}^T$ where A is the symmetrical part of ρ:

$$A = \frac{\rho + \rho^T}{2} = \begin{bmatrix} \rho_1 & \beta_1 & & & \\ \beta_1 & \rho_2 & \beta_2 & & \\ & \ddots & \ddots & \ddots & \\ & & \beta_{N-2} & \rho_{N-1} & \beta_{N-1} \\ & & & \beta_{N-1} & \rho_N \end{bmatrix} \tag{5.87}$$

where $\beta_i = (\rho_{i,i+1} + \rho_{i+1,i})/2$ for $i = 1, \cdots, N-1$. Physically P has to be non-negative for any prescribed current $\boldsymbol{J}$. This is equivalent to require that A is positive semi-definite, which means that all the determinants of the principal submatrices of

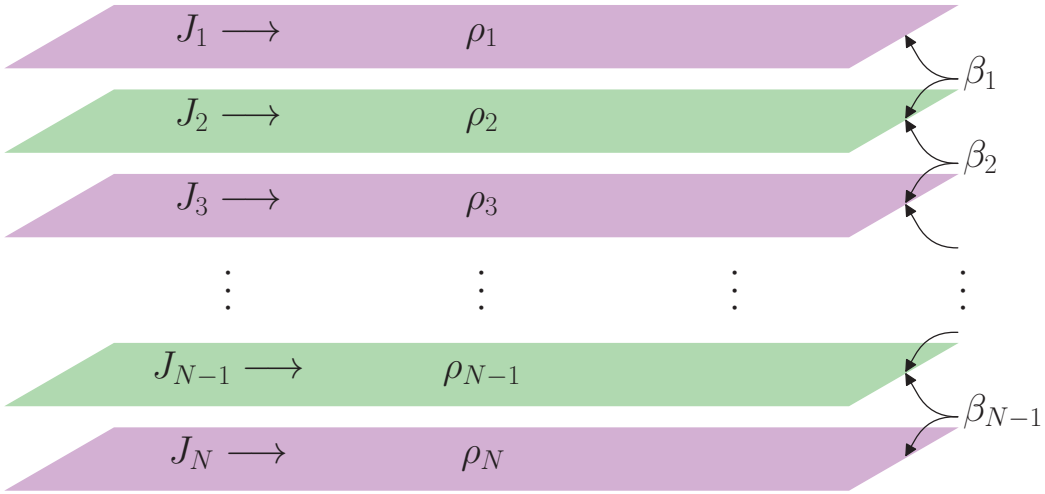

Fig. 5.6. Schematics of a multilayer system where ρ_i $(i = 1, \cdots, N)$ is the intralayer resistivity and β_i $(i = 1, \cdots, N-1)$ represents the interlayer coupling. J_i is the electrical current flow in layer i. The arrows $\longrightarrow$ represent the current direction. The layers are colored alternatively only for distinction.

A are non-negative [186]. Particularly for tridiagonal matrix A, it requires $\rho_i \geq 0$ $(i = 1, 2, \cdots, N)$ and

$$A(m,n) = \begin{vmatrix} \rho_m & \beta_m & & & & \\ \beta_m & \rho_{m+1} & \beta_{m+1} & & & \\ & \ddots & \ddots & \ddots & & \\ & & \beta_{n-2} & \rho_{n-1} & \beta_{n-1} & \\ & & & \beta_{n-1} & \rho_n & \end{vmatrix} \geq 0 \tag{5.88}$$

for $1 \leq m < n \leq N$. We will have a set of inequality equations. For example, if $n = m + 1$, we have

$$\beta_i^2 \leq \rho_i \rho_{i+1}, \quad i = 1, \cdots, N-1. \tag{5.89}$$

If one layer is superconducting, e.g., $\rho_i = 0$ $(i \neq 1, N)$, then $\beta_i = \beta_{i-1} = 0$ from Eq. 5.89. It implies $\rho_{i-1,i} = \rho_{i+1,i} = 0$ because $\rho_{i,i-1} = \rho_{i,i+1} = 0$ since no electrical field can be established in the superconducting layer (for $i = 1, N$, $\rho_1 = 0 \Rightarrow \beta_1 = 0 \Rightarrow \rho_{1,2} = 0$ and $\rho_N = 0 \Rightarrow \beta_{N-1} = 0 \Rightarrow \rho_{N,N-1} = 0$). This is a counterintuitive result, because the Cooper pairs in superconductors are charged and supposed to be able to drag the charge carriers in the adjacent normal conductor. Nevertheless, our prediction is consistent with the experiments [173–176].

The existence of superconducting layers will divide the matrix A into block diagonal form with some blocks being 1×1 matrices (non-negative entries) and the others being tridiagonal matrices that have the same appearance as A. The vanishing of β_i will also divide A into block diagonal form. Note that the diagonal entries of these block tridiagonal matrices are positive. The positive semi-definiteness of A requires each block to be positive semi-definite. Thus, it is convenient to assume that $\rho_i > 0$ $(i = 1, \cdots, N)$ and $\beta_i \neq 0$ $(i = 1, \cdots, N-1)$ when discussing the positive semi-definiteness of A or calculating the total power consumption.

We apply the above discussions to thermoelectric transport in multilayer systems and investigate how the Coulomb drag between layers affect the thermoelectric transport. The traditional thermoelectric modules consist of p- and n-type cylinders mounted between cold and hot plates [144]. The cylinders are connected in a convenient way so that the electrical current flow in the n-type is opposite to that in the p-type cylinders. Mapping this idea to the multilayer systems discussed above, the current flow direction in adjacent layers should be opposite, which is referred as the counterflow transport where the current in layer i is $J_i = (-1)^i J$. The total power consumption now becomes

$$P_c = J^2 \left(\sum_{i=1}^{N} \rho_i - 2 \sum_{i=1}^{N-1} \beta_i \right). \tag{5.90}$$

Apparently we have $\sum_{i=1}^{N}\rho_i \geq 2\sum_{i=1}^{N-1}\beta_i$. If the equality holds, we will achieve dissipationless counter flow transport so that $P_c = 0$ for any magnitude of J. This is useful for enhancing the thermoelectric efficiency as we will see later. For a two-layer system, it is obvious that we need to have $\beta_1 = \dfrac{\rho_1+\rho_2}{2} \leq \sqrt{\rho_1\rho_2}$. The only solution is $\beta_1 = \rho_1 = \rho_2$, as shown in the previous sections. Before providing a general form of A that permits such dissipationless transport, we work out this problem for a three-layer system. The positive semi-definiteness of A then requires that all of its leading principle minors are non-negative, i.e.,

$$\begin{aligned} \rho_1 &\geq 0, \\ \rho_1\rho_2 - \beta_1^2 &\geq 0, \\ \rho_1\rho_2\rho_3 - \rho_3\beta_1^2 - \rho_1\beta_2^2 &\geq 0. \end{aligned} \tag{5.91}$$

We now need to find out the maximum value of $\beta_1+\beta_2$ for β_1 and β_2 inside the ellipse of $\rho_3\beta_1^2 + \rho_1\beta_2^2 \leq \rho_1\rho_2\rho_3$. Apparently the maximum of $\beta_1+\beta_2$ can be found on the boundary of this ellipse. Thus,

$$\max\{\beta_1+\beta_2\} = \max\left\{\beta_1 + \sqrt{\rho_2\rho_3 - \frac{\rho_3}{\rho_1}\beta_1^2}\right\} \tag{5.92}$$

where $\beta_1^2 \leq \rho_1\rho_2$. This maximum can be achieved for $\beta_1^2 = \dfrac{\rho_1^2\rho_2}{\rho_1+\rho_3}$ and $\beta_2^2 = \dfrac{\rho_3^2\rho_2}{\rho_1+\rho_3}$. Now $\max\{\beta_1+\beta_2\} = \sqrt{\rho_2(\rho_1+\rho_3)}$. Apparently $\rho_1+\rho_2+\rho_3 \geq 2\max\{\beta_1+\beta_2\}$ and the equality holds if $\rho_2 = \rho_1+\rho_3$. Thus $\beta_1 = \rho_1$ and $\beta_2 = \rho_3$.

This result can be extended to systems with arbitrary number of layers. Given any $\rho_i > 0$ $(i = 1, \cdots, N-1)$, we need to have

$$\beta_i = \sum_{j=1}^{i}(-1)^{j+i}\rho_i > 0, \quad i = 1, \cdots, N-1 \tag{5.93}$$

and $\rho_N = \beta_{N-1} > 0$. Note that the positiveness of β_i imposes restrictions on the values of ρ_i. Now A is diagonal dominant, since $\rho_{i+1} = \beta_i + \beta_{i+1}$ $(i = 1, \cdots, N-2)$,

$\rho_1 = \beta_1$ and $\rho_N = \beta_{N-1}$. Therefore A is positive semi-definite. [187, p. 18] We can verify that indeed $P_c = 0$ for any magnitude of J. There might exist solutions other than Eq. 5.93 that allow $P_c = 0$, which is out of the scope of our discussions.

5.6 Summary

In this chapter, we offer a phenomenological theory of how the presence of strong Coulomb drag is affecting the (thermoelectric) transport properties of double layer systems. For Coulomb drag resistivities that satisfying some specific relations, the (counterflow) transport in double (multiple) layer system could be dissipationless. The Drude-like model provides some insights about the conditions on the relaxation time of Coulomb scattering between carriers in different layers. The Boltzmann formalism is employed to study the Coulomb drag conductivity and its reciprocal relations. It is pointed out that the existence of interlayer excitons might break the reciprocal relations. Our theory is also generalized to a generic multi-layer system.

6. SUMMARY AND FUTURE DIRECTIONS

6.1 Summary

The majority part of this thesis is focusing on energy and carrier transport in few graphene-based systems, from both theoretical[1] and experimental[2] perspectives. Along this journey, some generic theoretical models[3] that can be applied beyond graphene-based systems are developed with the inspiration input from it. Some important results are listed below:

- [*Simulation*] Molecular dynamics (MD) simulation is a convenient tool to study especially the lattice thermal transport for nano-scale systems. In general, two thermal baths at two different temperatures (with temperature difference ΔT) are attached at two ends of GNRs of various size (length, width), chirality (armchair, zigzag) and shape (rectangular, triangular, trapezoidal). Heat current flowing across the GNRs can be calculated. According to the magnitude of ΔT (relative to the average temperature of the GNRs), our MD simulations roughly fall in two different categories of small (linear response regime) and large (nonlinear response regime) ΔT. For the cases of small ΔT, the thermal conductivity/conductance can be defined/calculated and is dependent on a single temperature, i.e., the average temperature of the GNRs.

[1]Labeled by [*Simulation*]
[2]Labeled by [*Experiment*]
[3]Labeled by [*Theory*]

1. Though the calculated thermal conductivity for GNRs with symmetrical shape is close to the reported experimental value, its exact value is of limited predictive power; instead, its relative variation with respect to quantities such as temperature and GNR chirality is of physical significance.
2. For symmetrical GNRs, thermal conductivity increases with temperature and the zigzag GNRs have larger thermal conductivity than the armchair GNRs.
3. For asymmetrical GNRs, thermal rectification is shown to exist and the thermal rectification factor can be engineered by the shape of GNRs.
4. The existence of vacancies/defects can dramatically modify the thermal conductivity of GNRs and weaken the effect of thermal rectification.
5. The edge passionation by hydrogen can reduce the thermal conductivity of GNRs.
6. Isotopic doping can modulate the thermal conductivity of GNRs.

For the case of large ΔT, the thermal conductivity/conductance cannot be properly defined/calculated since the heat current depends on not only ΔT but also the average temperature, which provides two degrees of freedom to control the thermal transport in GNRs.

1. NDTR is shown to exist for large ΔT, when one end is at a constant temperature while the temperature at the other end is reduced (there is no NDTR if this temperature is increased).
2. NDTR is weakened when the length of the GNRs is increased, and eventually it will disappear for very long GNRs, which leads us to the development of the theory on the existence of NDTR in 1D diffusive thermal transport.

- [*Theory*] For a 1D system that the thermal transport can be described by Fourier law with or without abrupt junctions, we predict that the existence of the junctions is necessary to observe NDTR.

- [*Experiment*] Thermal conductivity measurement using 3ω method is performed in synthesized graphene composite. It is shown that the thermal conductivity depends on the volume percent of graphene filling.

- [*Experiment*] Graphene double layer samples are fabricated using home-built flake transfer stage and Coulomb drag and counterflow thermoelectric transport measurements are performed.

    1. Coulomb drag resistivity between two bilayer-graphene layers is measured, with correct sign of the drag resistivity vs. carrier types in both layers. A simple model is proposed to explain the observation that the out-of-phase component of the drag resistivity is linearly dependent on the frequency of the drive current.

    2. Counterflow thermoelectric transport measurement shows that the measured counterflow Seebeck coefficient is the difference between the Seebeck coefficients of the individual graphene layer, with negligible influence from the small Coulomb drag resistivity. The intralayer resistivities are measured simultaneously with the counterflow Seebeck coefficient, so that the counterflow power factor can be calculated, and it can reach $\sim$ 700 μW/K^{2}cm at room temperature.

- [*Theory*] A theory for the transport in double layer systems is developed to study the effect of (esp. strong) Coulomb drag on the (particularly counterflow) carrier transport.

1. An upper limit on the amplitude of Coulomb drag resistivity is derived based on a simple model.
2. It is shown that for double layer systems with large Coulomb drag resistivity, the (counterflow) transport could be dissipationless, promising high ZT modules.
3. A Drude-like model is used to study the transport in double layer systems with Coulomb drag (described by a drag relaxation time constant) in presence. It is found that the existence of interlayer Coulomb scattering tends to enhance the counterflow effective resistivity.
4. The Boltzmann transport formalism is employed to include excitons on equal footing with free carriers. It is found that the reciprocal relations for the Coulomb drag conductivities could be break down if excitons are present. The existence of interlayer excitons is necessary for achieving low counterflow effective resistivity.
5. The theory is also generalized to generic multi-layer systems.

A few other *experiments* (one in Appendix F and the rest are not included in the thesis) are conducted in graphene-based systems. Although these are not completed, it is worth to mention them here:

- Graphene flakes are transferred onto SiO_2/silicon substrate that has arrays of nanoscale craters or trenches fabricated by focused ion beam. Metal contacts to graphene is then made and the laser shock technique is used to make graphene conform with the surface profile of the substrate to induce various periodic patterns of stress in graphene. Raman spectroscopy is used to shown that the stress in graphene is created. Preliminary (low temperature) transport measurement results shown no new features for graphene with periodic stress.

- Graphene/BN/$NbSe_2$ samples are fabricated using the home-built flake transfer stage. A representative sample and the transport measurement data are shown in Appendix F. Superconducting transition is observed in the thin $NbSe_2$ substrate. Though graphene is only 5 nm away from $NbSe_2$, no significant feature is seen in the colormaps of the longitudinal/transverse resistivity of graphene vs. temperature and magnetic field. Coulomb drag between graphene and $NbSe_2$ is performed, but the drag resistivity vs. graphene carrier density (gate voltages) exhibit very similar features for temperatures below and above the superconducting transition temperature of $NbSe_2$.

- Graphene/$NbSe_2$ samples are fabricated where graphene layer is between $NbSe_2$ and SiO_2/silicon. For most of them graphene is electrically isolated from $NbSe_2$. No induced superconductivity in graphene is observed.

- Graphene/BN/NbN samples with very thin BN layer ($< 3nm$) are fabricated, where the NbN layer is deposited using suppering tool. The superconducting transition in the NbN layer is observed. The expected screening of the magnetic field for temperatures below the superconducting transition temperature does not exist for all our samples.

- The transfer of chemical vapor deposition (CVD) graphene grown on copper deposited on sapphire substrates onto the SiO_2/silicon substrate using wafer bonder and metal film peeling is practiced to fabricate large size graphene Hall bar samples for developing graphene-based quantum Hall resistance standards.

6.2 Future directions

The *simulation/theory/experiment* in this thesis and the incomplete *experiments* lead us to the following aspects that we can focus on in future:

- Finding out the appropriate thermal contact junctions [112] that can present NDTR effect, either using MD simulation tools/models or doing experiments, will be interesting towards building modules of NDTR that has not been realized yet.

- It is interesting to study the thermoelectric transport in graphene composite [165, 188, 189]. As the percolation behaviour of electrical transport is already demonstrated in graphene composites, the thermoelectric transport of these composites could also exhibit strong dependence on the graphene filling fractions. For thermal transport in graphene based composites, phonons are the dominate energy carrier (usually the matrix material is insulative) and they can go through the matrix material and get scattered by the graphene-matrix (or graphene-graphene) interface. Therefore the thermal transport doesnt exhibit percolation behaviour, as shown in Fig. (3.2 b). We thus could have different dependence of the electrical conductivity, thermal conductivity and Seebeck coefficient on graphene filling fractions, potentially useful for engineering thermoelectric properties. This will enable us to optimize the graphene filling fraction to obtain optimal thermoelectric figure of merit.

- It is appealing to build a double layer system with strong Coulomb drag resistivity which is approaching the limit predicted in Chapter 5. Then our theory of dissipationless (counterflow) transport and thermoelectric transport with en-

hanced ZT can be tested. Graphene double layer is promising since strong Coulomb drag is already observed [130].

- Our theory also predicts that the Coulomb drag resistivity between a superconducting layer and a non-superconducting layer should be zero when the drive current is in the superconducting layer. This can be tested using materials with high superconducting transition temperature, since the materials ($NbSe_2$, Nb and NbN) we have tried have low transition temperature at which the Coulomb drag resistivity is usually too small.

- It is predicted that graphene on type-II superconducting materials with controllable vortices could be a new platform for realizing quantum computing [190, 191]. Our experiments on this subject will be continued in future.

- We briefly involve in strain engineering in graphene. We hope we can make progress on building resistance standard using strain-induced magnetic field without using real magnet [192].

- Graphene with superconducting contact [193–195].

REFERENCES

REFERENCES

[1] A. K. Geim and K. S. Novoselov, "The rise of graphene," *Nature Materials*, vol. 6, no. 3, pp. 183–191, 2007.

[2] K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, Y. Zhang, S. V. Dubonos, I. V. Grigorieva, and A. A. Firsov, "Electric field effect in atomically thin carbon films," *Science*, vol. 306, no. 5696, pp. 666–669, 2004.

[3] G. Ruan, Z. Sun, Z. Peng, and J. M. Tour, "Growth of graphene from food, insects, and waste," *ACS Nano*, vol. 5, no. 9, pp. 7601–7607, 2011.

[4] X. Li, W. Cai, J. An, S. Kim, J. Nah, D. Yang, R. Piner, A. Velamakanni, I. Jung, E. Tutuc, S. K. Banerjee, L. Colombo, and R. S. Ruoff, "Large-area synthesis of high-quality and uniform graphene films on copper foils," *Science*, vol. 324, no. 5932, pp. 1312–1314, 2009.

[5] C. Berger, Z. Song, X. Li, X. Wu, N. Brown, C. Naud, D. Mayou, T. Li, J. Hass, A. N. Marchenkov *et al.*, "Electronic confinement and coherence in patterned epitaxial graphene," *Science*, vol. 312, no. 5777, pp. 1191–1196, 2006.

[6] Y.-M. Lin, C. Dimitrakopoulos, K. A. Jenkins, D. B. Farmer, H.-Y. Chiu, A. Grill, and P. Avouris, "100-ghz transistors from wafer-scale epitaxial graphene," *Science*, vol. 327, no. 5966, pp. 662–662, 2010.

[7] K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, M. I. Katsnelson, I. V. Grigorieva, S. V. Dubonos, and A. A. Firsov, "Two-dimensional gas of massless dirac fermions in graphene," *Nature*, vol. 438, p. 197, 2005.

[8] Y. Zhang, Y.-W. Tan, H. L. Stormer, and P. Kim, "Experimental observation of the quantum hall effect and berrys phase in graphene," *Nature*, vol. 438, no. 7065, p. 201, 2005.

[9] A. C. Neto, F. Guinea, N. Peres, K. S. Novoselov, and A. K. Geim, "The electronic properties of graphene," *Reviews of Modern Physics*, vol. 81, no. 1, p. 109, 2009.

[10] N. M. R. Peres, "*Colloquium*: The transport properties of graphene: An introduction," *Reviews of Modern Physics*, vol. 82, pp. 2673–2700, Sep 2010.

[11] S. Das Sarma, S. Adam, E. H. Hwang, and E. Rossi, "Electronic transport in two-dimensional graphene," *Reviews of Modern Physics*, vol. 83, pp. 407–470, May 2011.

[12] J. Drut, "Graphene: from materials science to particle physics." in *Proceedings of The XXVIII International Symposium on Lattice Field Theory*, vol. 1, 2010, p. 6.

[13] S. Stankovich, D. A. Dikin, G. H. B. Dommett, K. M. Kohlhaas, E. J. Zimney, E. A. Stach, R. D. Piner, S. T. Nguyen, and R. S. Ruoff, "Graphene-based composite materials." *Nature*, vol. 442, no. 7100, pp. 282–286, 2006.

[14] C. Lee, X. Wei, J. W. Kysar, and J. Hone, "Measurement of the elastic properties and intrinsic strength of monolayer graphene," *Science*, vol. 321, no. 5887, pp. 385–388, 2008.

[15] A. R. Ranjbartoreh, B. Wang, X. Shen, and G. Wang, "Advanced mechanical properties of graphene paper," *Journal of Applied Physics*, vol. 109, no. 1, p. 014306, 2011.

[16] K. Nakada, M. Fujita, G. Dresselhaus, and M. S. Dresselhaus, "Edge state in graphene ribbons: Nanometer size effect and edge shape dependence," *Physical Review B*, vol. 54, no. 24, p. 17954, 1996.

[17] Y.-W. Son, M. L. Cohen, and S. G. Louie, "Energy gaps in graphene nanoribbons," *Physical Review Letters*, vol. 97, no. 21, p. 216803, 2006.

[18] M. Y. Han, B. Özyilmaz, Y. Zhang, and P. Kim, "Energy band-gap engineering of graphene nanoribbons," *Physical Review Letters*, vol. 98, no. 20, p. 206805, 2007.

[19] Z. Chen, Y.-M. Lin, M. J. Rooks, and P. Avouris, "Graphene nano-ribbon electronics," *Physica E: Low-dimensional Systems and Nanostructures*, vol. 40, no. 2, pp. 228–232, 2007.

[20] L. A. Ponomarenko, A. K. Geim, A. A. Zhukov, R. Jalil, S. V. Morozov, K. S. Novoselov, I. V. Grigorieva, E. H. Hill, V. V. Cheianov, V. I. Fal'ko, K. Watanabe, T. Taniguchi, and R. V. Gorbachev, "Tunable metalinsulator transition in double-layer graphene heterostructures," *Nature Physics*, vol. 7, no. 12, pp. 958–961, Dec. 2011.

[21] J. Hu, X. Ruan, and Y. P. Chen, "Thermal conductivity and thermal rectification in graphene nanoribbons: a molecular dynamics study," *Nano letters*, vol. 9, no. 7, pp. 2730–2735, 2009.

[22] J. Hu, X. Ruan, Z. Jiang, and Y. P. Chen, "Molecular dynamics calculation of thermal conductivity of graphene nanoribbons," in *AIP Conference Proceedings*, 2009, pp. 135–138.

[23] J. Hu, X. Ruan, and Y. P. Chen, "Molecular dynamics simulation of melting and vacancy movement in graphene nanoribbons," in *Proceedings of the 30th International Thermal Conductivity Conference and the 18th International Thermal Expansion Symposium*, 2009, p. 750.

[24] J. Hu, S. Schiffli, A. Vallabhaneni, X. Ruan, and Y. P. Chen, "Tuning the thermal conductivity of graphene nanoribbons by edge passivation and isotope engineering: A molecular dynamics study," *Applied Physics Letters*, vol. 97, no. 13, p. 133107, 2010.

[25] J. Hu, Y. Wang, A. Vallabhaneni, X. Ruan, and Y. P. Chen, "Nonlinear thermal transport and negative differential thermal conductance in graphene nanoribbons," *Applied Physics Letters*, vol. 99, p. 113101, 2011.

[26] J. Hu, X. Ruan, and Y. P. Chen, "Molecular dynamics study of thermal rectification in graphene nanoribbons," *International Journal of Thermophysics*, vol. 33, no. 6, pp. 986–991, 2012.

[27] Y. Wang, A. Vallabhaneni, J. Hu, B. Qiu, Y. P. Chen, and X. Ruan, "Phonon lateral confinement enables thermal rectification in asymmetric single-material nanostructures," *Nano Letters*, vol. 14, no. 2, pp. 592–596, 2014.

[28] A. K. Vallabhaneni, B. Qiu, J. Hu, Y. P. Chen, A. K. Roy, and X. Ruan, "Interfacial thermal conductance limit and thermal rectification across vertical carbon nanotube/graphene nanoribbon-silicon interfaces," *Journal of Applied Physics*, vol. 113, no. 6, p. 064311, 2013.

[29] D. G. Cahill, "Thermal conductivity measurement from 30 to 750 k: the 3 omega method," *Review of Scientific Instruments*, vol. 61, no. 2, pp. 802–808, 1990.

[30] C. Dean, A. Young, I. Meric, C. Lee, L. Wang, S. Sorgenfrei, K. Watanabe, T. Taniguchi, P. Kim, and K. Shepard, "Boron nitride substrates for high-quality graphene electronics," *Nature Nanotechnology*, vol. 5, no. 10, pp. 722–726, 2010.

[31] C. R. Dean, A. F. Young, P. Cadden-Zimansky, L. Wang, H. Ren, K. Watanabe, T. Taniguchi, P. Kim, J. Hone, and K. L. Shepard, "Multicomponent fractional quantum hall effect in graphene," *Nature Physics*, vol. 7, no. 9, pp. 693–696, 2011.

[32] D. W. Brenner, "Empirical potential for hydrocarbons for use in simulating the chemical vapor deposition of diamond films," *Physical Review B*, vol. 42, no. 15, p. 9458, 1990.

[33] D. W. Brenner, O. A. Shenderova, J. A. Harrison, S. J. Stuart, B. Ni, and S. B. Sinnott, "A second-generation reactive empirical bond order (REBO) potential energy expression for hydrocarbons," *Journal of Physics: Condensed Matter*, vol. 14, no. 4, p. 783, 2002.

[34] S. Nosé, "A unified formulation of the constant temperature molecular dynamics methods," *The Journal of Chemical Physics*, vol. 81, no. 1, pp. 511–519, 1984.

[35] W. G. Hoover, "Canonical dynamics: equilibrium phase-space distributions," *Physical Review A*, vol. 31, no. 3, p. 1695, 1985.

[36] L. Jiao, B. Fan, X. Xian, Z. Wu, J. Zhang, and Z. Liu, "Creation of nanostructures with poly(methyl methacrylate)-mediated nanotransfer printing," *Journal of the American Chemical Society*, vol. 130, no. 38, pp. 12 612–12 613, Sep. 2008.

[37] L. Jiao, X. Xian, Z. Wu, J. Zhang, and Z. Liu, "Selective positioning and integration of individual single-walled carbon nanotubes," *Nano Letters*, vol. 9, no. 1, pp. 205–209, Jan. 2009.

[38] T. Taychatanapat, K. Watanabe, T. Taniguchi, and P. Jarillo-Herrero, "Quantum hall effect and landau-level crossing of dirac fermions in trilayer graphene," *Nature Physics*, vol. 7, no. 8, pp. 621–625, 2011.

[39] J. Eisenstein and A. MacDonald, "Bose–einstein condensation of excitons in bilayer electron systems," *Nature*, vol. 432, no. 7018, pp. 691–694, 2004.

[40] H. Min, R. Bistritzer, J.-J. Su, and A. MacDonald, "Room-temperature superfluidity in graphene bilayers," *Physical Review B*, vol. 78, no. 12, p. 121401, 2008.

[41] Y. E. Lozovik and A. Sokolik, "Electron-hole pair condensation in a graphene bilayer," *JETP Letters*, vol. 87, no. 1, pp. 55–59, 2008.

[42] M. Y. Kharitonov and K. B. Efetov, "Electron screening and excitonic condensation in double-layer graphene systems," *Physical Review B*, vol. 78, no. 24, p. 241401, 2008.

[43] C. Zhang and Y. Joglekar, "Excitonic condensation of massless fermions in graphene bilayers," *Physical Review B*, vol. 77, no. 23, p. 233405, 2008.

[44] Y. E. Lozovik, S. Ogarkov, and A. Sokolik, "Condensation of electron-hole pairs in a two-layer graphene system: Correlation effects," *Physical Review B*, vol. 86, no. 4, p. 045429, 2012.

[45] D. Abergel, R. Sensarma, and S. D. Sarma, "Density fluctuation effects on the exciton condensate in double-layer graphene," *Physical Review B*, vol. 86, no. 16, p. 161412, 2012.

[46] I. Sodemann, D. Pesin, and A. MacDonald, "Interaction-enhanced coherence between two-dimensional dirac layers," *Physical Review B*, vol. 85, no. 19, p. 195136, 2012.

[47] Y. F. Suprunenko, V. Cheianov, and V. I. Fal'ko, "Phases of the excitonic condensate in two-layer graphene," *Physical Review B*, vol. 86, no. 15, p. 155405, 2012.

[48] A. Perali, D. Neilson, and A. R. Hamilton, "High-temperature superfluidity in double-bilayer graphene," *Physical Review Letters*, vol. 110, no. 14, p. 146803, 2013.

[49] D. Abergel, M. Rodriguez-Vega, E. Rossi, and S. D. Sarma, "Interlayer excitonic superfluidity in graphene," *Physical Review B*, vol. 88, no. 23, p. 235402, 2013.

[50] D. Nandi, A. Finck, J. Eisenstein, L. Pfeiffer, and K. West, "Exciton condensation and perfect coulomb drag," *Nature*, vol. 488, no. 7412, pp. 481–484, 2012.

[51] A. A. Balandin, "Thermal properties of graphene and nanostructured carbon materials," *Nature Materials*, vol. 10, no. 8, pp. 569–581, 2011.

[52] C. Chang, D. Okawa, A. Majumdar, and A. Zettl, "Solid-state thermal rectifier," *Science*, vol. 314, no. 5802, pp. 1121–1124, 2006.

[53] G. Wu and B. Li, "Thermal rectifiers from deformed carbon nanohorns," *Journal of Physics: Condensed Matter*, vol. 20, no. 17, p. 175211, 2008.

[54] G. Wu and B. Li, "Thermal rectification in carbon nanotube intramolecular junctions: Molecular dynamics calculations," *Physical Review B*, vol. 76, no. 8, p. 085424, 2007.

[55] B. Li, L. Wang, and G. Casati, "Thermal diode: Rectification of heat flux," *Physical Review Letters*, vol. 93, no. 18, p. 184301, 2004.

[56] B. Li, L. Wang, and G. Casati, "Negative differential thermal resistance and thermal transistor," *Applied Physics Letters*, vol. 88, no. 14, p. 143501, 2006.

[57] L. Wang and B. Li, "Thermal logic gates: Computation with phonons," *Physical Review Letters*, vol. 99, no. 17, p. 177208, 2007.

[58] L. Wang and B. Li, "Thermal memory: a storage of phononic information," *Physical review letters*, vol. 101, no. 26, p. 267203, 2008.

[59] P. H. Hnenberger, "Thermostat algorithms for molecular dynamics simulations," *Advances in Polymer Science*, vol. 173, no. 130, p. 105, 2005.

[60] L. Falkovsky, "Symmetry constraints on phonon dispersion in graphene," *Physics Letters A*, vol. 372, p. 5189, 2008.

[61] A. A. Balandin, S. Ghosh, W. Bao, I. Calizo, D. Teweldebrhan, F. Miao, and C. N. Lau, "Superior thermal conductivity of single-layer graphene," *Nano letters*, vol. 8, no. 3, pp. 902–907, 2008.

[62] W. Cai, A. L. Moore, Y. Zhu, X. Li, S. Chen, L. Shi, and R. S. Ruoff, "Thermal transport in suspended and supported monolayer graphene grown by chemical vapor deposition," *Nano Letters*, vol. 10, no. 5, pp. 1645–1651, 2010.

[63] C. Faugeras, B. Faugeras, M. Orlita, M. Potemski, R. R. Nair, and A. Geim, "Thermal conductivity of graphene in corbino membrane geometry," *ACS Nano*, vol. 4, no. 4, pp. 1889–1892, 2010.

[64] L. A. Jauregui, Y. Yue, A. N. Sidorov, J. Hu, Q. Yu, G. Lopez, R. Jalilian, D. K. Benjamin, D. A. Delkd, W. Wu *et al.*, "Thermal transport in graphene nanostructures: Experiments and simulations," *ECS Transactions*, vol. 28, no. 5, pp. 73–83, 2010.

[65] J. H. Seol, I. Jo, A. L. Moore, L. Lindsay, Z. H. Aitken, M. T. Pettes, X. Li, Z. Yao, R. Huang, D. Broido, N. Mingo, R. S. Ruoff, and L. Shi, "Two-Dimensional Phonon Transport in Supported Graphene," *Science*, vol. 328, no. 5975, p. 213, 2010.

[66] J.-W. Jiang, J.-S. Wang, and B. Li, "Thermal conductance of graphene and dimerite," *Physical Review B*, vol. 79, p. 205418, May 2009.

[67] G. Zhang and B. Li, "Thermal conductivity of nanotubes revisited: Effects of chirality, isotope impurity, tube length, and temperature," *The Journal of Chemical Physics*, vol. 123, no. 11, p. 114714, 2005.

[68] T. Yamamoto, S. Watanabe, and K. Watanabe, "Universal features of quantized thermal conductance of carbon nanotubes," *Physical Review Letters*, vol. 92, p. 075502, Feb 2004.

[69] S. Ghosh, I. Calizo, D. Teweldebrhan, E. Pokatilov, D. Nika, A. Balandin, W. Bao, F. Miao, and C. N. Lau, "Extremely high thermal conductivity of graphene: Prospects for thermal management applications in nanoelectronic circuits," *Applied Physics Letters*, vol. 92, no. 15, pp. 151 911–151 911, 2008.

[70] D. L. Nika, S. Ghosh, E. P. Pokatilov, and A. A. Balandin, "Lattice thermal conductivity of graphene flakes: Comparison with bulk graphite," *Applied Physics Letters*, vol. 94, no. 20, p. 203103, 2009.

[71] D. L. Nika, E. P. Pokatilov, A. S. Askerov, and A. A. Balandin, "Phonon thermal conduction in graphene: Role of umklapp and edge roughness scattering," *Physical Review B*, vol. 79, p. 155413, Apr 2009.

[72] M. Peyrard, "The design of a thermal rectifier," *Europhysics Letters*, vol. 76, p. 49, 2006.

[73] C. Dames, "Solid-state thermal rectification with existing bulk materials," *Journal of Heat Transfer*, vol. 131, no. 6, p. 061301, 2009.

[74] A. K. Vallabhaneni, J. Hu, Y. P. Chen, and X. Ruan, "Thermal rectification in graphene and carbon nanotube systems using molecular dynamics simulations," in *ASME/JSME 2011 8th Thermal Engineering Joint Conference*. American Society of Mechanical Engineers, 2011, pp. T30 060–T30 060.

[75] K. Gunawardana, K. Mullen, J. Hu, Y. P. Chen, and X. Ruan, "Tunable thermal transport and thermal rectification in strained graphene nanoribbons," *Physical Review B*, vol. 85, no. 24, p. 245417, 2012.

[76] N. Yang, G. Zhang, and B. Li, "Thermal rectification in asymmetric graphene ribbons," *Applied Physics Letters*, vol. 95, no. 3, p. 033107, 2009.

[77] Y. Xu, X. Chen, J.-S. Wang, B.-L. Gu, and W. Duan, "Thermal transport in graphene junctions and quantum dots," *Physical Review B*, vol. 81, no. 19, p. 195425, 2010.

[78] J.-W. Jiang, J. Chen, J.-S. Wang, and B. Li, "Edge states induce boundary temperature jump in molecular dynamics simulation of heat conduction," *Phys. Rev. B*, vol. 80, no. 5, p. 052301, 2009.

[79] Y. Xu, X. Chen, B.-L. Gu, and W. Duan, "Intrinsic anisotropy of thermal conductance in graphene nanoribbons," *Applied Physics Letters*, vol. 95, no. 23, p. 233116, 2009.

[80] S. Park and R. S. Ruoff, "Chemical methods for the production of graphenes," *Nature Nanotechnology*, vol. 4, no. 4, pp. 217–224, 2009.

[81] D. Boukhvalov and M. Katsnelson, "Chemical functionalization of graphene," *Journal of Physics: Condensed Matter*, vol. 21, no. 34, p. 344205, 2009.

[82] Y. H. Lu, R. Q. Wu, L. Shen, M. Yang, Z. D. Sha, Y. Q. Cai, P. M. He, and Y. P. Feng, "Effects of edge passivation by hydrogen on electronic structure of armchair graphene nanoribbon and band gap engineering," *Applied Physics Letters*, vol. 94, no. 12, p. 122111, 2009.

[83] X. Li, W. Cai, L. Colombo, and R. S. Ruoff, "Evolution of graphene growth on ni and cu by carbon isotope labeling," *Nano Letters*, vol. 9, no. 12, pp. 4268–4272, 2009.

[84] W. J. Evans, L. Hu, and P. Keblinski, "Thermal conductivity of graphene ribbons from equilibrium molecular dynamics: Effect of ribbon width, edge roughness, and hydrogen termination," *Applied Physics Letters*, vol. 96, no. 20, p. 203112, 2010.

[85] N. W. Ashcroft and N. D. Mermin, *Solid State Physics*. Brooks Cole, 1976.

[86] F. Müller-Plathe, "A simple nonequilibrium molecular dynamics method for calculating the thermal conductivity," *The Journal of Chemical Physics*, vol. 106, no. 14, pp. 6082–6085, 1997.

[87] S. Plimpton, "Fast parallel algorithms for short-range molecular dynamics," *Journal of Computational Physics*, vol. 117, no. 1, pp. 1–19, 1995.

[88] G. Zhang and B. Li, "Thermal conductivity of nanotubes revisited: Effects of chirality, isotope impurity, tube length, and temperature," *The Journal of Chemical Physics*, vol. 123, no. 11, p. 114714, 2005.

[89] G. Stoltz, N. Mingo, and F. Mauri, "Reducing the thermal conductivity of carbon nanotubes below the random isotope limit," *Physical Review B*, vol. 80, no. 11, p. 113408, 2009.

[90] C. Chang, A. Fennimore, A. Afanasiev, D. Okawa, T. Ikuno, H. Garcia, D. Li, A. Majumdar, and A. Zettl, "Isotope effect on the thermal conductivity of boron nitride nanotubes," *Physical Review Letters*, vol. 97, no. 8, p. 085901, 2006.

[91] D. A. Stewart, I. Savic, and N. Mingo, "First-principles calculation of the isotope effect on boron nitride nanotube thermal conductivity," *Nano Letters*, vol. 9, no. 1, pp. 81–84, 2008.

[92] N. Yang, G. Zhang, and B. Li, "Ultralow thermal conductivity of isotope-doped silicon nanowires," *Nano Letters*, vol. 8, no. 1, pp. 276–280, 2008.

[93] W. Kim, J. Zide, A. Gossard, D. Klenov, S. Stemmer, A. Shakouri, and A. Majumdar, "Thermal conductivity reduction and thermoelectric figure of merit increase by embedding nanoparticles in crystalline semiconductors," *Physical Review Letters*, vol. 96, no. 4, p. 045901, 2006.

[94] J.-W. Jiang, J. Lan, J.-S. Wang, and B. Li, "Isotopic effects on the thermal conductivity of graphene nanoribbons: Localization mechanism," *Journal of Applied Physics*, vol. 107, no. 5, p. 054314, 2010.

[95] J. Chen, G. Zhang, and B. Li, "Tunable thermal conductivity of si 1-x ge x nanowires," *Applied Physics Letters*, vol. 95, no. 7, pp. 073 117–073 117, 2009.

[96] T. Ouyang, Y. Chen, K. Yang, and J. Zhong, "Thermal transport of isotopic-superlattice graphene nanoribbons with zigzag edge," *Europhysics Letters*, vol. 88, no. 2, p. 28002, 2009.

[97] W.-R. Zhong, P. Yang, B.-Q. Ai, Z.-G. Shao, and B. Hu, "Negative differential thermal resistance induced by ballistic transport," *Physical Review E*, vol. 79, no. 5, p. 050103, 2009.

[98] D. He, S. Buyukdagli, and B. Hu, "Origin of negative differential thermal resistance in a chain of two weakly coupled nonlinear lattices," *Physical Review B*, vol. 80, no. 10, p. 104302, 2009.

[99] E. Pereira, "Graded anharmonic crystals as genuine thermal diodes: Analytical description of rectification and negative differential thermal resistance," *Physical Review E*, vol. 82, no. 4, p. 040101, 2010.

[100] D. He, B.-Q. Ai, H.-K. Chan, and B. Hu, "Heat conduction in the nonlinear response regime: Scaling, boundary jumps, and negative differential thermal resistance," *Physical Review E*, vol. 81, no. 4, p. 041131, 2010.

[101] D. Segal, "Heat flow in nonlinear molecular junctions: Master equation analysis," *Physical Review B*, vol. 73, no. 20, p. 205415, 2006.

[102] D. M.-T. Kuo and Y.-C. Chang, "Thermoelectric effects of multiple quantum dot junctions in the nonlinear response regime," *Japanese Journal of Applied Physics*, vol. 49, p. 064301, 2010.

[103] N. Yang, N. Li, L. Wang, and B. Li, "Thermal rectification and negative differential thermal resistance in lattices with mass gradient," *Physical Review B*, vol. 76, no. 2, p. 020301, 2007.

[104] Z.-G. Shao, L. Yang, H.-K. Chan, and B. Hu, "Transition from the exhibition to the nonexhibition of negative differential thermal resistance in the two-segment frenkel-kontorova model," *Physical Review E*, vol. 79, no. 6, p. 061119, 2009.

[105] B.-Q. Ai and B. Hu, "Heat conduction in deformable frenkel-kontorova lattices: Thermal conductivity and negative differential thermal resistance," *Physical Review E*, vol. 83, no. 1, p. 011131, 2011.

[106] B.-Q. Ai, W.-R. Zhong, and B. Hu, "Double negative differential thermal resistance induced by nonlinear on-site potentials," *Physical Review E*, vol. 83, no. 5, p. 052102, 2011.

[107] Z.-G. Shao and L. Yang, "Relationship between negative differential thermal resistance and ballistic transport," *Europhysics Letters*, vol. 94, no. 3, p. 34004, 2011.

[108] L. Zhu, C. R. Otey, and S. Fan, "Negative differential thermal conductance through vacuum," *Applied Physics Letters*, vol. 100, p. 044104, 2012.

[109] L.-A. Wu, C. X. Yu, and D. Segal, "Nonlinear quantum heat transfer in hybrid structures: Sufficient conditions for thermal rectification," *Physical Review B*, vol. 80, p. 041103, 2009.

[110] L. Esaki, "New phenomenon in narrow germanium p-n junctions," *Physical Review*, vol. 109, no. 2, pp. 603–604, Jan 1958.

[111] Z. Huang and Z. Tang, "Evaluation of momentum conservation influence in non-equilibrium molecular dynamics methods to compute thermal conductivity," *Physica B: Condensed Matter*, vol. 373, no. 2, pp. 291–296, 2006.

[112] J. Hu and Y. P. Chen, "Existence of negative differential thermal conductance in one-dimensional diffusive thermal transport," *Physical Review E*, vol. 87, p. 062104, 2013.

[113] E. Swartz and R. Pohl, "Thermal boundary resistance," *Reviews of Modern Physics*, vol. 61, pp. 605–668, Jul 1989.

[114] K. M. F. Shahil and A. A. Balandin, "Graphenemultilayer graphene nanocomposites as highly efficient thermal interface materials," *Nano Letters*, vol. 12, no. 2, pp. 861–867, 2012.

[115] F. Yavari, H. R. Fard, K. Pashayi, M. A. Rafiee, A. Zamiri, Z. Yu, R. Ozisik, T. Borca-Tasciuc, and N. Koratkar, "Enhanced thermal conductivity in a nanostructured phase change composite due to low concentration graphene additives," *The Journal of Physical Chemistry C*, vol. 115, no. 17, pp. 8753–8758, 2011.

[116] W. S. Hummers and R. E. Offeman, "Preparation of graphitic oxide," *Journal of the American Chemical Society*, vol. 80, no. 6, pp. 1339–1339, 1958.

[117] D. C. Marcano, D. V. Kosynkin, J. M. Berlin, A. Sinitskii, Z. Sun, A. Slesarev, L. B. Alemany, W. Lu, and J. M. Tour, "Improved synthesis of graphene oxide," *ACS Nano*, vol. 4, no. 8, pp. 4806–4814, 2010.

[118] W. Park, J. Hu, L. A. Jauregui, X. Ruan, and Y. P. Chen, "Electrical and thermal conductivities of reduced graphene oxide/polystyrene composites," *Applied Physics Letters*, vol. 104, no. 11, p. 113101, 2014.

[119] S. Yu, P. Hing, and X. Hu, "Thermal conductivity of polystyrene-aluminum nitride composite," *Composites Part A: Applied Science and Manufacturing*, vol. 33, no. 2, pp. 289–292, 2002.

[120] R. Prasher, C.-P. Chiu, and R. Mahajan, "Thermal interface materials: a brief review of design characteristics and materials," *Electron Cooling*, vol. 8, p. 12, 2004.

[121] C. Dames and G. Chen, "1ω, 2ω, and 3ω methods for measurements of thermal properties," *Review of Scientific Instruments*, vol. 76, no. 12, p. 124902, 2005.

[122] P. E. Hopkins, B. Kaehr, L. M. Phinney, T. P. Koehler, A. M. Grillet, D. Dunphy, F. Garcia, and C. J. Brinker, "Measuring the thermal conductivity of porous, transparent sio[sub 2] films with time domain thermoreflectance," *Journal of Heat Transfer*, vol. 133, no. 6, p. 061601, 2011.

[123] R. G. Morris and J. G. Hust, "Thermal conductivity measurements of silicon from 30 °C to 425 °C," *Physical Review*, vol. 124, pp. 1426–1430, Dec 1961.

[124] W. Yu, H. Xie, X. Wang, and X. Wang, "Significant thermal conductivity enhancement for nanofluids containing graphene nanosheets," *Physics Letters A*, vol. 375, no. 10, pp. 1323–1328, 2011.

[125] T. Schwamb, B. R. Burg, N. C. Schirmer, and D. Poulikakos, "An electrical method for the measurement of the thermal and electrical conductivity of reduced graphene oxide nanostructures," *Nanotechnology*, vol. 20, no. 40, p. 405704, 2009.

[126] J. Chiu and P. Fair, "Determination of thermal conductivity by differential scanning calorimetry," *Thermochimica Acta*, vol. 34, no. 2, pp. 267–273, 1979.

[127] Y. P. Koh, G. B. McKenna, and S. L. Simon, "Calorimetric glass transition temperature and absolute heat capacity of polystyrene ultrathin films," *Journal of Polymer Science Part B: Polymer Physics*, vol. 44, no. 24, 2006.

[128] L. Britnell, R. V. Gorbachev, R. Jalil, B. D. Belle, F. Schedin, M. I. Katsnelson, L. Eaves, S. V. Morozov, A. S. Mayorov, and N. M. Peres, "Electron tunneling through ultrathin boron nitride crystalline barriers," *Nano Letters*, vol. 12, no. 3, pp. 1707–1710, 2012.

[129] A. Geim and I. Grigorieva, "Van der waals heterostructures," *Nature*, vol. 499, no. 7459, pp. 419–425, 2013.

[130] R. Gorbachev, A. Geim, M. Katsnelson, K. Novoselov, T. Tudorovskiy, I. Grigorieva, A. MacDonald, S. Morozov, K. Watanabe, and T. Taniguchi, "Strong coulomb drag and broken symmetry in double-layer graphene," *Nature Physics*, vol. 8, no. 12, pp. 896–901, 2012.

[131] L. Ponomarenko, A. Geim, A. Zhukov, R. Jalil, S. Morozov, K. Novoselov, I. Grigorieva, E. Hill, V. Cheianov, and V. FalKo, "Tunable metal-insulator transition in double-layer graphene heterostructures," *Nature Physics*, vol. 7, no. 12, pp. 958–961, 2011.

[132] T. Georgiou, R. Jalil, B. D. Belle, L. Britnell, R. V. Gorbachev, S. V. Morozov, Y.-J. Kim, A. Gholinia, S. J. Haigh, and O. Makarovsky, "Vertical field-effect transistor based on graphene-ws2 heterostructures for flexible and transparent electronics," *Nature Nanotechnology*, vol. 8, no. 2, pp. 100–103, 2013.

[133] L. Britnell, R. Gorbachev, A. Geim, L. Ponomarenko, A. Mishchenko, M. Greenaway, T. Fromhold, K. Novoselov, and L. Eaves, "Resonant tunnelling and negative differential conductance in graphene transistors," *Nature Communications*, vol. 4, p. 1794, 2013.

[134] K. Lee, B. Fallahazad, J. Xue, D. C. Dillen, K. Kim, T. Taniguchi, K. Watanabe, and E. Tutuc, "Chemical potential and quantum hall ferromagnetism in bilayer graphene," *Science*, vol. 345, no. 6192, pp. 58–61, 2014.

[135] M. J. Manfra, "Molecular beam epitaxy of ultra-high-quality algaas/gaas heterostructures: Enabling physics in low-dimensional electronic systems," *Annual Review of Condensed Matter Physics*, vol. 5, no. 1, pp. 347–373, 2014.

[136] C. Goupil, W. Seifert, K. Zabrocki, E. Müller, and G. J. Snyder, "Thermodynamics of thermoelectric phenomena and applications," *Entropy*, vol. 13, no. 8, pp. 1481–1517, 2011.

[137] D. M. Rowe, *Thermoelectrics handbook: macro to nano*. CRC press, 2005.

[138] S. Lin, P. Tong, B. Wang, J. Lin, Y. Huang, and Y. Sun, "Good thermoelectric performance in strongly correlated system sncco3 with antiperovskite structure," *Inorganic Chemistry*, vol. 53, no. 7, pp. 3709–3715, 2014.

[139] J. Clarke and S. Freake, "Superconducting fountain effect," *Physical Review Letters*, vol. 29, no. 9, p. 588, 1972.

[140] T. Karpiuk, B. Grémaud, C. Miniatura, and M. Gajda, "Superfluid fountain effect in a bose-einstein condensate," *Physical Review A*, vol. 86, no. 3, p. 033619, 2012.

[141] A. Rancon, C. Chin, and K. Levin, "Bosonic thermoelectric transport and breakdown of universalit," *arXiv:1311.0769*, 2013.

[142] S. Kim and E. Tutuc, "Coulomb drag and magnetotransport in graphene double layers," *Solid State Communications*, vol. 152, no. 15, pp. 1283–1288, 2012.

[143] J. Lux and L. Fritz, "Interaction-dominated transport and coulomb drag in bilayer graphene," *Physical Review B*, vol. 87, no. 7, p. 075423, 2013.

[144] G. J. Snyder and E. S. Toberer, "Complex thermoelectric materials," *Nature Materials*, vol. 7, no. 2, pp. 105–114, 2008.

[145] M. S. Dresselhaus, G. Chen, M. Y. Tang, R. Yang, H. Lee, D. Wang, Z. Ren, J.-P. Fleurial, and P. Gogna, "New directions for low-dimensional thermoelectric materials," *Advanced Materials*, vol. 19, no. 8, pp. 1043–1053, 2007.

[146] E. Pop, V. Varshney, and A. K. Roy, "Thermal properties of graphene: Fundamentals and applications," *Mrs Bulletin*, vol. 37, no. 12, pp. 1273–1281, 2012.

[147] J.-H. Chen, C. Jang, S. Xiao, M. Ishigami, and M. S. Fuhrer, "Intrinsic and extrinsic performance limits of graphene devices on sio2," *Nature Nanotechnology*, vol. 3, no. 4, pp. 206–209, 2008.

[148] N. Sankeshwar, S. Kubakaddi, and B. Mulimani, "Thermoelectric power in graphene," *Intech*, vol. 9, pp. 1–56, 2013.

[149] H. Sevinçli and G. Cuniberti, "Enhanced thermoelectric figure of merit in edge-disordered zigzag graphene nanoribbons," *Physical Review B*, vol. 81, no. 11, p. 113401, 2010.

[150] Y. Chen, T. Jayasekera, A. Calzolari, K. Kim, and M. B. Nardelli, "Thermoelectric properties of graphene nanoribbons, junctions and superlattices," *Journal of Physics: Condensed Matter*, vol. 22, no. 37, p. 372202, 2010.

[151] F. Mazzamuto, V. Hung Nguyen, Y. Apertet, C. Caër, C. Chassat, J. Saint-Martin, and P. Dollfus, "Enhanced thermoelectric properties in graphene nanoribbons by resonant tunneling of electrons," *Physical Review B*, vol. 83, p. 235426, Jun 2011.

[152] W. Huang, J.-S. Wang, and G. Liang, "Theoretical study on thermoelectric properties of kinked graphene nanoribbons," *Physical Review B*, vol. 84, no. 4, p. 045410, 2011.

[153] T. Gunst, T. Markussen, A.-P. Jauho, and M. Brandbyge, "Thermoelectric properties of finite graphene antidot lattices," *Physical Review B*, vol. 84, no. 15, p. 155449, 2011.

[154] L. Liang, E. Cruz-Silva, E. C. Girão, and V. Meunier, "Enhanced thermoelectric figure of merit in assembled graphene nanoribbons," *Physical Review B*, vol. 86, no. 11, p. 115438, 2012.

[155] H. Zheng, H. Liu, X. Tan, H. Lv, L. Pan, J. Shi, and X. Tang, "Enhanced thermoelectric performance of graphene nanoribbons," *Applied Physics Letters*, vol. 100, no. 9, p. 093104, 2012.

[156] Z.-X. Xie, L.-M. Tang, C.-N. Pan, K.-M. Li, K.-Q. Chen, and W. Duan, "Enhancement of thermoelectric properties in graphene nanoribbons modulated with stub structures," *Applied Physics Letters*, vol. 100, no. 7, p. 073105, 2012.

[157] G. Kliros and P. Divari, "Thermoelectric properties of gated graphene ribbons in the ballistic regime," *Superlattices and Microstructures*, vol. 52, no. 2, pp. 221–233, 2012.

[158] K. Yang, Y. Chen, R. D'Agosta, Y. Xie, J. Zhong, and A. Rubio, "Enhanced thermoelectric properties in hybrid graphene/boron nitride nanoribbons," *Physical Review B*, vol. 86, p. 045425, Jul 2012.

[159] H. Karamitaheri, N. Neophytou, M. Pourfath, R. Faez, and H. Kosina, "Engineering enhanced thermoelectric properties in zigzag graphene nanoribbons," *Journal of Applied Physics*, vol. 111, no. 5, p. 054501, 2012.

[160] P.-H. Chang and B. K. Nikolić, "Edge currents and nanopore arrays in zigzag and chiral graphene nanoribbons as a route toward high-z t thermoelectrics," *Physical Review B*, vol. 86, no. 4, p. 041406, 2012.

[161] C.-N. Pan, Z.-X. Xie, L.-M. Tang, and K.-Q. Chen, "Ballistic thermoelectric properties in graphene-nanoribbon-based heterojunctions," *Applied Physics Letters*, vol. 101, no. 10, p. 103115, 2012.

[162] H. Sevinçli, C. Sevik, T. Çağın, and G. Cuniberti, "A bottom-up route to enhance thermoelectric figures of merit in graphene nanoribbons," *Scientific Reports*, vol. 3, 2013.

[163] P. S. E. Yeo, M. B. Sullivan, K. P. Loh, and C. K. Gan, "First-principles study of the thermoelectric properties of strained graphene nanoribbons," *Journal of Materials Chemistry A*, vol. 1, no. 36, pp. 10 762–10 767, 2013.

[164] B. Feng, J. Xie, G. Cao, T. Zhu, and X. Zhao, "Enhanced thermoelectric properties of p-type $CoSb_3$/graphene nanocomposite," *Journal of Materials Chemistry A*, vol. 1, no. 42, pp. 13 111–13 119, 2013.

[165] H. Chen, C. Yang, H. Liu, G. Zhang, D. Wan, and F. Huang, "Thermoelectric properties of $CuInTe_2$/graphene composites," *CrystEngComm*, vol. 15, no. 34, pp. 6648–6651, 2013.

[166] J. P. Small, K. M. Perez, and P. Kim, "Modulation of thermoelectric power of individual carbon nanotubes," *Physical Review Letters*, vol. 91, p. 256801, Dec 2003.

[167] W. S. Leong, H. Gong, and J. T. Thong, "Low-contact-resistance graphene devices with nickel-etched-graphene contacts," *ACS Nano*, vol. 8, no. 1, pp. 994–1001, 2013.

[168] S. V. Ovsyannikov, V. V. Shchennikov, G. V. Vorontsov, A. Y. Manakov, A. Y. Likhacheva, and V. A. Kulbachinskii, "Giant improvement of thermoelectric power factor of bi 2 te 3 under pressure," *Journal of Applied Physics*, vol. 104, no. 5, pp. 053 713–053 713, 2008.

[169] K. Yazawa and A. Shakouri, "Cost-efficiency trade-off and the design of thermoelectric power generators," *Environmental Science & Technology*, vol. 45, no. 17, pp. 7548–7553, 2011.

[170] K. Wu, L. Rademaker, and J. Zaanen, "Bilayer excitons in two-dimensional nanostructures for greatly enhanced thermoelectric efficiency," *Physical Review Applied*, vol. 2, p. 054013, Nov 2014.

[171] Y. Chen, "Surface excitonic thermoelectric devices," 7 2012, US Patent App. 13/312, 986.

[172] A. Croxall, K. D. Gupta, C. Nicoll, M. Thangaraj, H. Beere, I. Farrer, D. Ritchie, and M. Pepper, "Anomalous coulomb drag in electron-hole bilayers," *Physical Review Letters*, vol. 101, no. 24, p. 246801, 2008.

[173] N. Giordano and J. Monnier, "Cross-talk effects in superconductor–insulator–normal-metal trilayers," *Physical Review B*, vol. 50, no. 13, p. 9363, 1994.

[174] X. Huang, G. Bazàn, and G. H. Bernstein, "Observation of supercurrent drag between normal metal and superconducting films," *Physical Review Letters*, vol. 74, no. 20, p. 4051, 1995.

[175] L. Farina, K. Lewis, C. Kurdak, S. Ghosh, and P. Bhattacharya, "Anomalous drag coupling between a thin superconducting film and a two-dimensional electron gas caused by radio frequency interference," *Physical Review B*, vol. 70, no. 15, p. 153302, 2004.

[176] L. Farina, K. Lewis, Ç. Kurdak, S. Ghosh, S. Krishna, and P. Bhattacharya, "Drag coupling between a thin al film and a two-dimensional electron gas near the superconducting transition," *Physica E: Low-dimensional Systems and Nanostructures*, vol. 22, no. 1, pp. 341–344, 2004.

[177] F. J. DiSalvo, "Thermoelectric cooling and power generation," *Science*, vol. 285, no. 5428, pp. 703–706, 1999.

[178] E. L. Hazlett, L.-C. Ha, and C. Chin, "Anomalous thermoelectric transport in two-dimensional bose gas," *arXiv:1306.4018*, 2013.

[179] F. Laliberté, J. Chang, N. Doiron-Leyraud, E. Hassinger, R. Daou, M. Rondeau, B. Ramshaw, R. Liang, D. Bonn, and W. Hardy, "Fermi-surface reconstruction by stripe order in cuprate superconductors," *Nature Communications*, vol. 2, p. 432, 2011.

[180] J. Allen and H. Jones, "New phenomena connected with heat flow in helium ii," *Nature*, vol. 141, no. 3562, pp. 243–244, 1938.

[181] J. Eisenstein, "Exciton condensation in bilayer quantum hall systems," *Annual Review of Condensed Matter Physics*, vol. 5, no. 1, pp. 159–181, 2014.

[182] G. Vignale and A. MacDonald, "Drag in paired electron-hole layers," *Physical Review Letters*, vol. 76, no. 15, p. 2786, 1996.

[183] Q. Chen, J. Stajic, S. Tan, and K. Levin, "Bcs–bec crossover: From high temperature superconductors to ultracold superfluids," *Physics Reports*, vol. 412, no. 1, pp. 1–88, 2005.

[184] A. Alkauskas, K. Flensberg, B. Y.-K. Hu, and A.-P. Jauho, "Sign reversal of drag in bilayer systems with in-plane periodic potential modulation," *Physical Review B*, vol. 66, p. 201304, Nov 2002.

[185] J. M. Ziman, *Electrons and Phonons: The Theory of Transport Phenomena in Solids*. Oxford University Press, 2001.

[186] K. Swamy, "On sylvester's criterion for positive-semidefinite matrices," *IEEE Transactions on Automatic Control*, vol. 18, no. 3, pp. 306–306, 1973.

[187] A. Greenbaum, *Iterative methods for solving linear systems.* SIAM, 1997, vol. 17.

[188] Y. Du, S. Z. Shen, W. Yang, R. Donelson, K. Cai, and P. S. Casey, "Simultaneous increase in conductivity and seebeck coefficient in a polyaniline/graphene nanosheets thermoelectric nanocomposite," *Synthetic Metals*, vol. 161, no. 23, pp. 2688–2692, 2012.

[189] M. Culebras, C. M. Gómez, and A. Cantarero, "Review on polymers for thermoelectric applications," *Materials*, vol. 7, no. 9, pp. 6701–6732, 2014.

[190] G. Rosenberg, B. Seradjeh, C. Weeks, and M. Franz, "Creation and manipulation of anyons in a layered superconductor˘two-dimensional electron gas system," *Physical Review B*, vol. 79, p. 205102, May 2009.

[191] C. Weeks, G. Rosenberg, B. Seradjeh, and M. Franz, "Anyons in a weakly interacting system," *Nature Physics*, vol. 3, no. 11, pp. 796–801, 2007.

[192] N. N. Klimov, S. Jung, S. Zhu, T. Li, C. A. Wright, S. D. Solares, D. B. Newell, N. B. Zhitenev, and J. A. Stroscio, "Electromechanical properties of graphene drumheads," *Science*, vol. 336, no. 6088, pp. 1557–1561, 2012.

[193] H. B. Heersche, P. Jarillo-Herrero, J. B. Oostinga, L. M. Vandersypen, and A. F. Morpurgo, "Bipolar supercurrent in graphene," *Nature*, vol. 446, no. 7131, pp. 56–59, 2007.

[194] P. Rickhaus, M. Weiss, L. Marot, and C. Schonenberger, "Quantum hall effect in graphene with superconducting electrodes," *Nano Letters*, vol. 12, no. 4, pp. 1942–1945, 2012.

[195] J.-H. Choi, G.-H. Lee, S. Park, D. Jeong, J.-O. Lee, H.-S. Sim, Y.-J. Doh, and H.-J. Lee, "Complete gate control of supercurrent in graphene p–n junctions," *Nature Communications*, vol. 4, 2013.

[196] A. Savvatimskiy, "Measurements of the melting point of graphite and the properties of liquid carbon (a review for 1963–2003)," *Carbon*, vol. 43, no. 6, pp. 1115–1142, 2005.

[197] G.-D. Lee, C. Wang, E. Yoon, N.-M. Hwang, D.-Y. Kim, and K. Ho, "Diffusion, coalescence, and reconstruction of vacancy defects in graphene layers," *Physical Review Letters*, vol. 95, no. 20, p. 205501, 2005.

[198] A. Y. Serov, Z.-Y. Ong, and E. Pop, "Effect of grain boundaries on thermal transport in graphene," *Applied Physics Letters*, vol. 102, no. 3, p. 033104, 2013.

[199] A. Bagri, S.-P. Kim, R. S. Ruoff, and V. B. Shenoy, "Thermal transport across twin grain boundaries in polycrystalline graphene from nonequilibrium molecular dynamics simulations," *Nano letters*, vol. 11, no. 9, pp. 3917–3921, 2011.

[200] O. V. Yazyev and Y. P. Chen, "Polycrystalline graphene and other two-dimensional materials," *Nature Nanotechnology*, vol. 9, pp. 755–767, 2014.

[201] F. Forster, A. Molina-Sanchez, S. Engels, A. Epping, K. Watanabe, T. Taniguchi, L. Wirtz, and C. Stampfer, "Dielectric screening of the kohn anomaly of graphene on hexagonal boron nitride," *Physical Review B*, vol. 88, p. 085419, Aug 2013.

[202] A. Parzygnat, K. K. Lee, Y. Avniel, and S. G. Johnson, "Sufficient conditions for two-dimensional localization by arbitrarily weak defects in periodic potentials with band gaps," *Physical Review B*, vol. 81, no. 15, p. 155324, 2010.

APPENDICES

A. MELTING OF GNR AND THERMAL TRANSPORT ACROSS GRAIN BOUNDARY USING MD SIMULATION

A.1 Melting and vacancy movements

We have used MD simulations to study the melting point and vacancy movement in a rectangular graphene nanoribbon[1]. The melting point of the graphene nanoribbon extracted from the numerical simulation is $\sim$ 3400 K. We also found that two separated vacancies at high temperature (e.g., $\sim$ 3000 K, below the melting point) can eventually form a stable Stone-Wales-like defect at the edge of the nanoribbon.

First, we study the melting of a rectangular GNR (2 nm $\times$ 3.5 nm). Fig. A.1 shows the snapshots of the GNR at t = 1 ns for four different specified temperatures (T_s) from 3100 K to 3400 K. Atoms denoted by squares (atoms at four side of the GNR) are fixed and the rest of the atoms (denoted by circles) are put in the Nos-Hoover thermostats at T_s indicated in the subtitles of Fig. A.1. The time t = 1 ns at which the snapshots are taken is large enough for the GNR to reach the thermal equilibrium. There are no lattice defects (except for bond length deformation) seen at T_s = 3100 K and T_s = 3200 K, while there are a few defects at T_s = 3300 K. However, at T_s = 3400 K, there are a few atoms escaped from the GNR. More atoms escape at larger simulation time, whereas it is stable at low temperature. We interpret this as the signature that the GNR has started to melt. The inset in the dashed blue rectangle in Fig. A.1d is the main part of the GNR zoomed in. Therefore, the melting point of

[1]This appendix is extracted from [23].

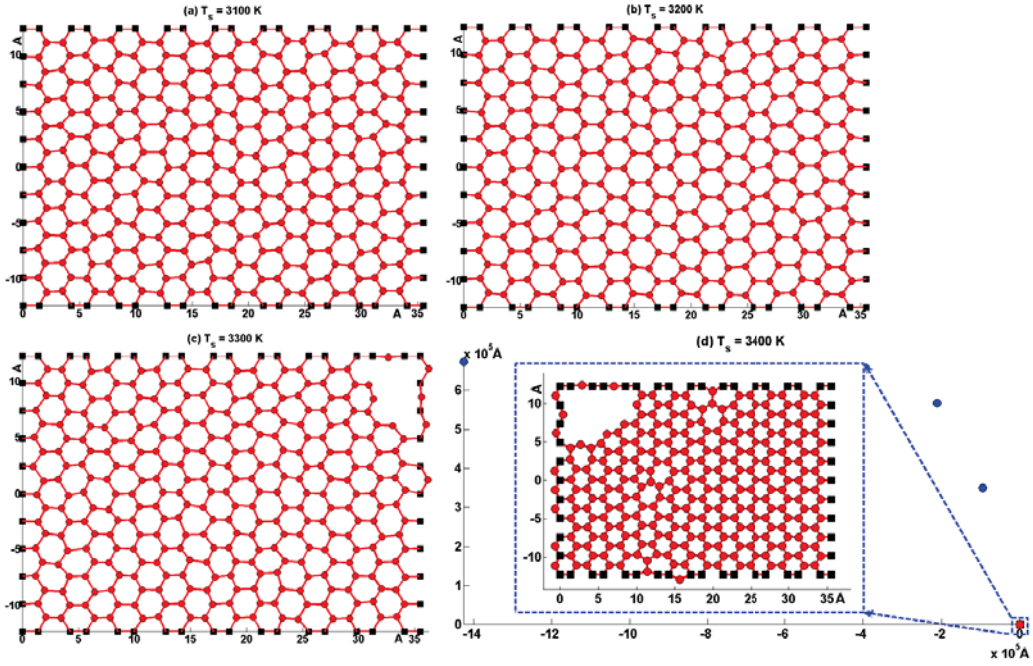

Fig. A.1. Snapshots of a square GNR at four different specified temperatures at t=1 ns. Note the different scale of the axis in (d). The solid blue circles denote the atoms escaped from the GNR.

GNR is extracted to be $\sim$ 3400 K, slightly lower than the melting point of graphite ($\sim$ 4000 K) [196].

In Fig. A.2 we plot the instant temperature T_i (measuring the average kinetic energy of atoms denoted by circles) of the GNR in Fig. A.1 at specified thermostat temperature T_s. When T_s is below 3300 K, the oscillation amplitude of T_i gradually decreases and T_i stabilizes around T_s. On the other hand, when T_s is above 3400 K, the oscillation amplitude of T_i increases and eventually saturates and T_i never stabilizes within the range of simulation time. This is consistent with the interpretation that the melting point of the GNR is $\sim$3400 K from Fig. A.1.

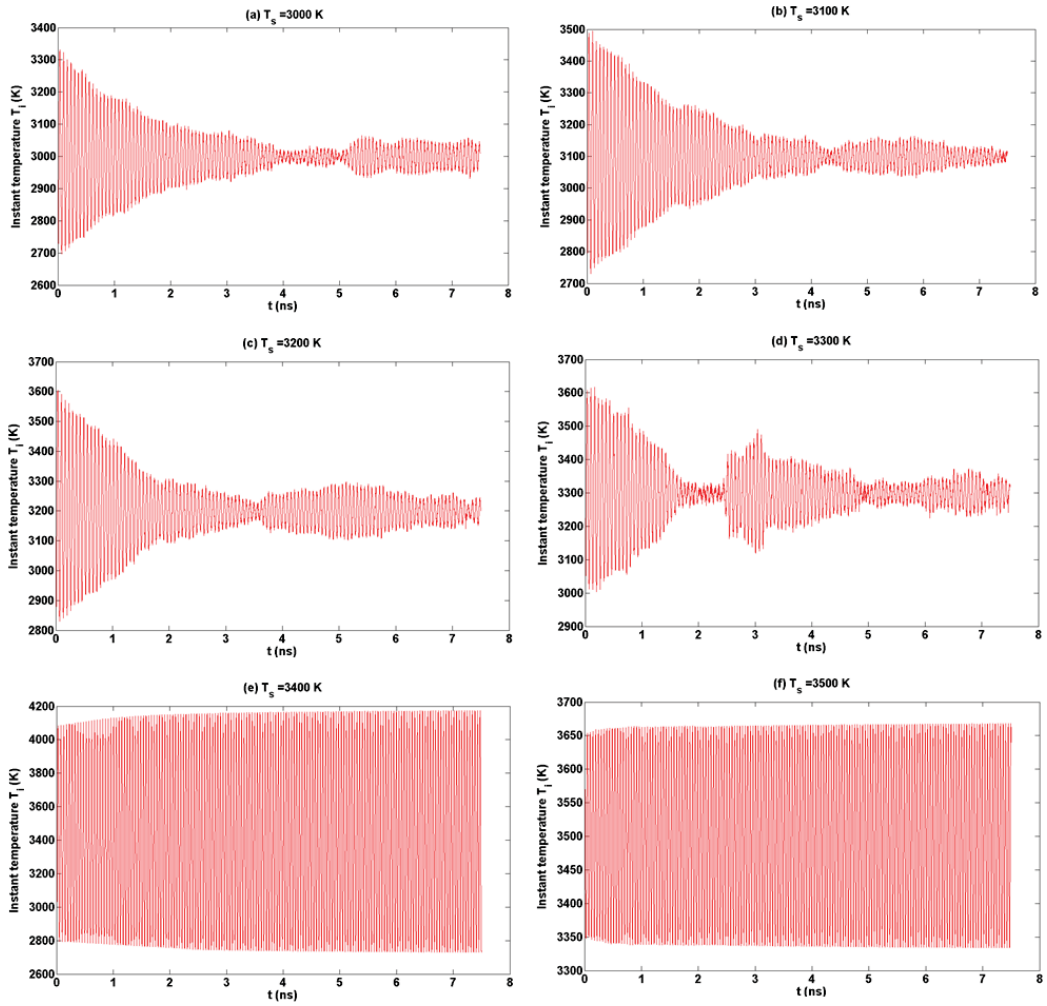

Fig. A.2. Oscillation of the instant temperature in the MD simulation of the GNR in Fig. A.1 at different specified temperatures.

Finally, in Fig. A.3 we show the movement of two vacancies in the same GNR as in Fig. A.1 at 3000 K, lower than the melting point ($\sim$ 3400 K) obtained from our simulation. At t=5 ps (Fig. A.3a), these two vacancies are separated. By t = 50 ps, one vacancy has diffused to the edge and the other has evolved into a 5-5 defect (not

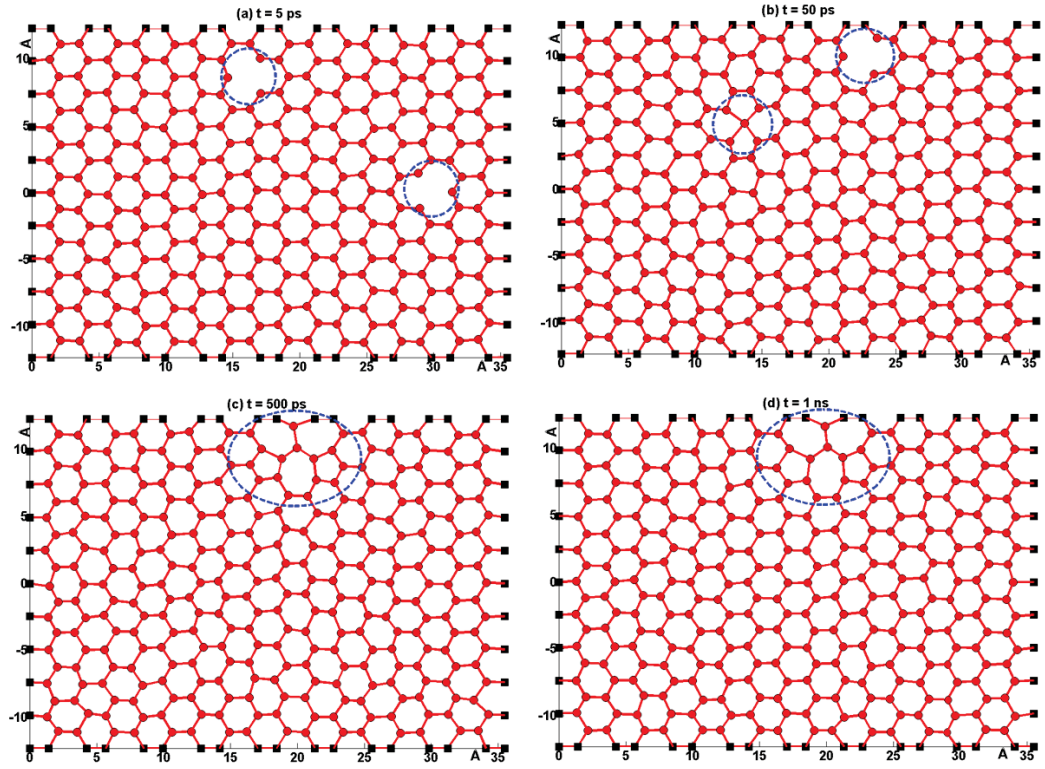

Fig. A.3. Vacancy movement at T_s=3000 K.

stable). They eventually form the stable 5-5-7-7-7 Stone-Wales-like defect near the edge of the GNR at long simulation time (0.5 1 ns). Relevant vacancies and defects in Fig. A.3 are indicated by the dashed circles/ellipses. A previous calculation based on density functional theory has obtained similar results [197].

A.2 Thermal transport across grain boundary

The structure of the grain boundary is shown in Fig. A.4. It is clear that the grain boundary structure is stable after performing structure relaxation at room temperature, using LAMMPS package [87].

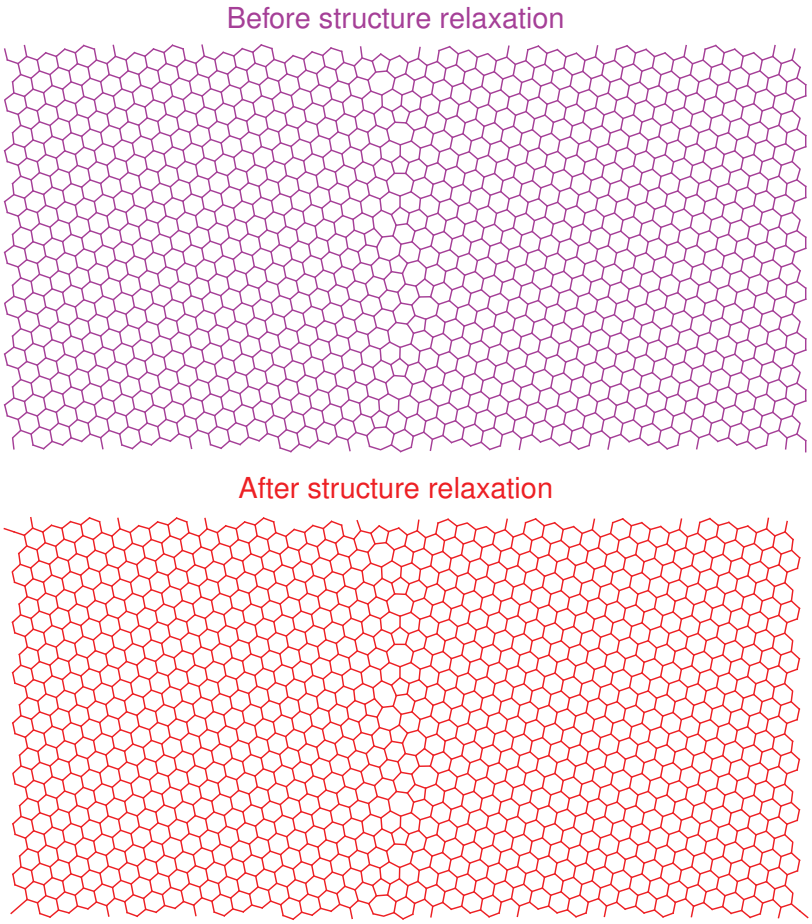

Fig. A.4. Structure of GNR with graphene boundary at the center before (top) and after (bottom) structure relaxation.

The method of velocity scaling [111] is used to create temperature gradient along the GNR. In this method, a heat current of 0.5 eV/fs is applied between two ends to generate nonequilibrium temperature distribution shown in Fig. A.5. The red region indicates the temperature drop across the grain boundary. The cyan regions are used to calculate the thermal conductivity of the region away from the grain boundary. The fitted thermal boundary conductance is about $\sim 10^{10}$ W/K-m^2, which agrees with the other simulations [198, 199]. The extracted thermal conductivity is about ~ 33 W/m-K.

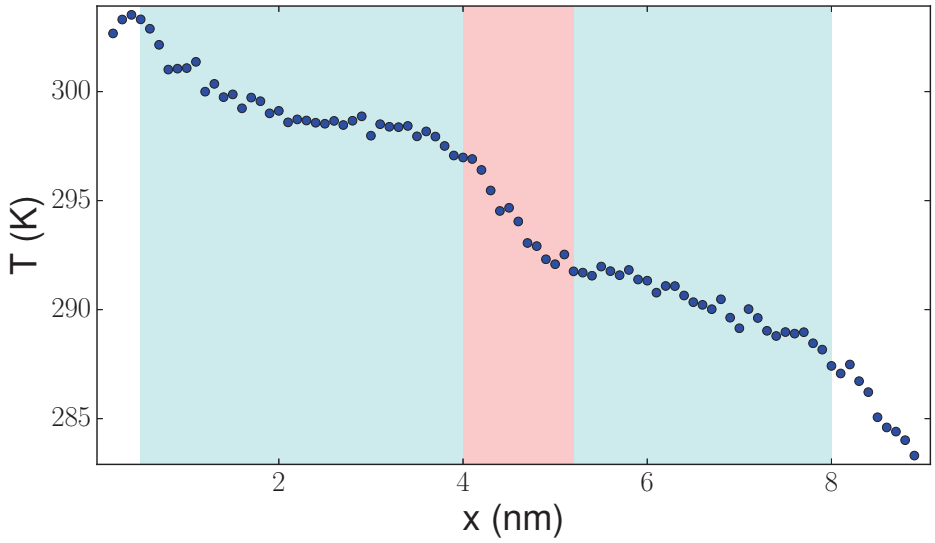

Fig. A.5. Temperature profile along the horizontal axis of the GNR in Fig. A.4.

Grain boundaries introduce additional scatterings for phonon transport, leading to an effective reduction of thermal conductivities. It those boundaries are transparent to electron transport, this will provide a route to engineering thermoelectric properties of graphene [200].

B. FLAKE TRANSFER SYSTEM AND PROCEDURE

Flake transfer refers to the procedure of placing flakes (typically 2D flakes of few to few tens of micrometer in size) to a desired location on a specified substrate. The dry transfer procedure we have used to fabricate multi-layer samples is shown in Fig. B.1, and it is summarized as the following:

(a) Coat poly(methyl methacrylate) (PMMA) layer on SiO_2/Si substrate and subsequently coat polyvinyl alcohol (PVA) layer on top.

(b) Exfoliate target flake onto the as-prepared substrate in (a), and attach adhesive tape on four edges of the substrate.

(c) Separate PMMA/PVA layer from SiO_2/Si substrate by grabbing the tape and slowly pulling the film off.

(d) Suspend the PMMA/PVA layer over a frame using the tape, and flip the frame to have the side with flakes facing downwards.

(e) Search for the flakes under microscope, and glue a small copper washer around the desired flake using PMMA, so that the flake is sitting roughly in the center of the washer. Cut the PMMA/PVA layer along the outer edge of the washer.

(f) Use the flake transfer stage to align the flake on the PMMA/PVA layer suspended over the washer with another desired flake on the substrate.

(g) Approach the washer to the substrate and make good contact by heating the substrate to around 120 ℃.

(h) Place the substrate with washer inside acetone bath to remove the PMMA layer and finish the flake transfer.

For our home-built system for flake transfer shown in Fig. B.2, the major components include sample chamber, welded bellow, XYZ translation stage, control valves, pump, microscope with long working distance lenses (100x and 50x are particularly useful) and temperature controller. Washers sitting on PMMA/PVA layer (2 inch by 2 inch suspended over a glass frame) are shown in Fig. B.2c, corresponding to Fig. B.1e. Since multiple usable flakes can be found, we usually take note for their locations first and then place the washers according to the quality (e.g., size and uniformity) and location of the flakes, e.g., if the distance between two flakes is about the radius of the washer, the one with lower quality will be covered by the edge of the washer and will not be used in order to expose the one with higher quality in the center of the washer.

The first demon of the transfer stage does not have a vacuum chamber, and the transferring of the graphene flake onto boron nitride often ends up with numerous bubbles at the graphene/boron nitride interface. At that time this issue of bubble is reported, but the mechanism of its occurrence is not clear. We then built the transfer stage shown in Fig. B.2a with a vacuum chamber so that the flake transfer can be operated in a vacuum environment and hoped that the bubbles would be gone. However, the bubbles remain even if the transfer is done in vacuum. We later learned that the occurrence of the bubbles is mainly due to the strain mismatch between substrate and the PMMA/PVA film. Nonetheless, building such vacuum environment is not a waste of effort; it dramatically reduces the rate of miss-alignment during flake contact.

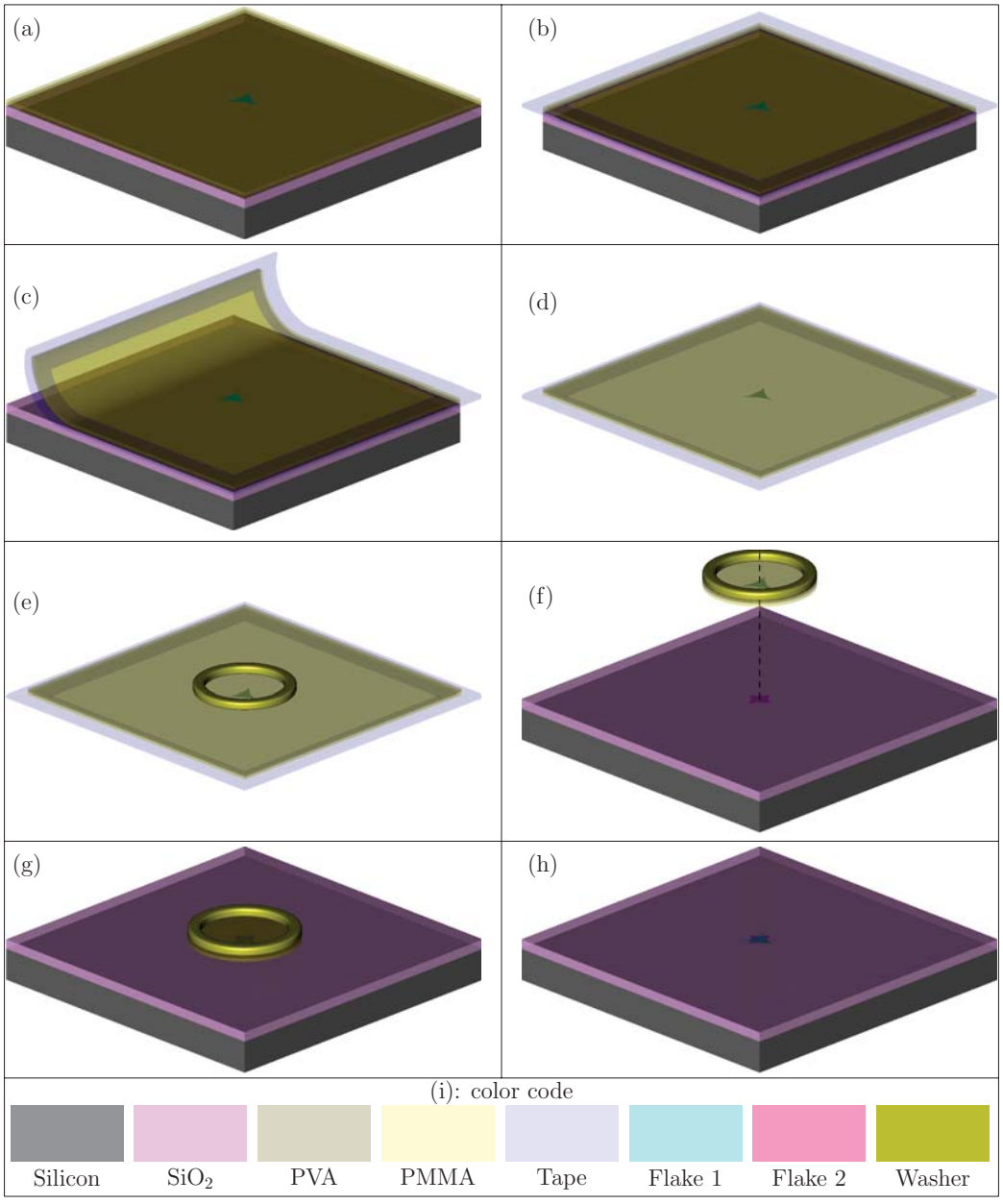

Fig. B.1. Flake transfer procedure (a-h) and its color code (i).

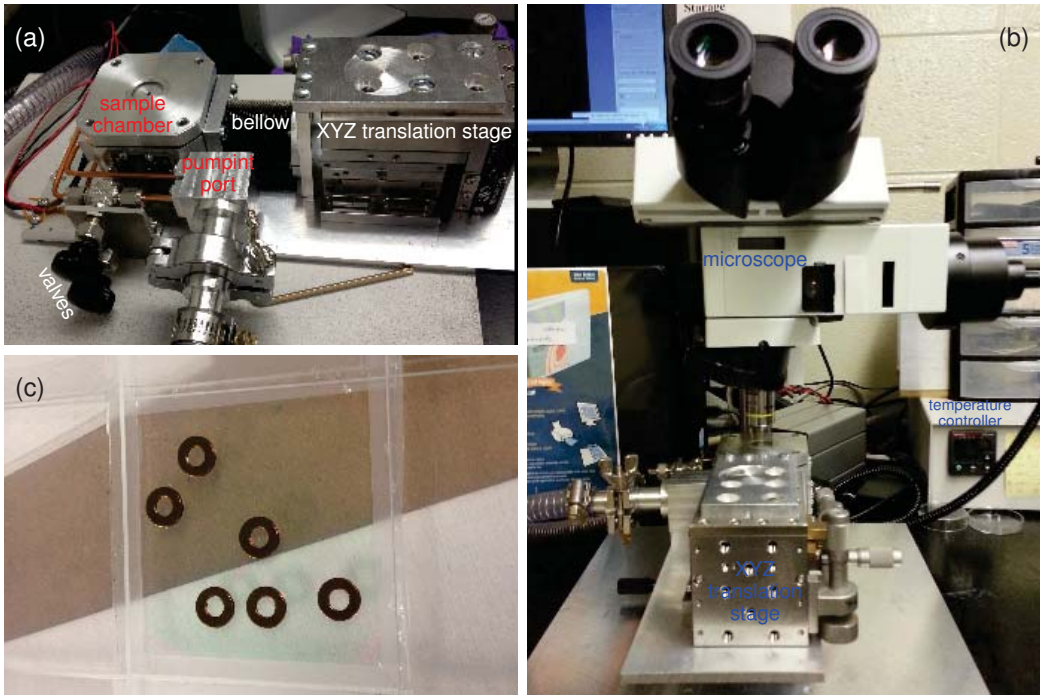

Fig. B.2. Flake transfer system: (a) flake transfer stage with some major components labeled, (b) stage mounted onto microscope and connected to a temperature controller and (c) the brass washer attached onto the PVA/PMMA film carrying the target flakes.

The operation of the transfer stage is shown in Fig. B.3. There are two chambers in the stage shown in Fig. B.3a. The silicon wafer is mounted on top of the metal arm below which is an embedded heater and temperature sensor to control the temperature of the substrate. The washer carrying the PMMA/PVA film on which the flake resides is glued on the metal separation between the top and bottom chambers. There is a hole on the metal separation similar to the size of the inner diameter of the washer in order to be able to see the silicon wafer through the glass window and the PMMA/PVA film. Two chambers are then pumped down to the base pressure of

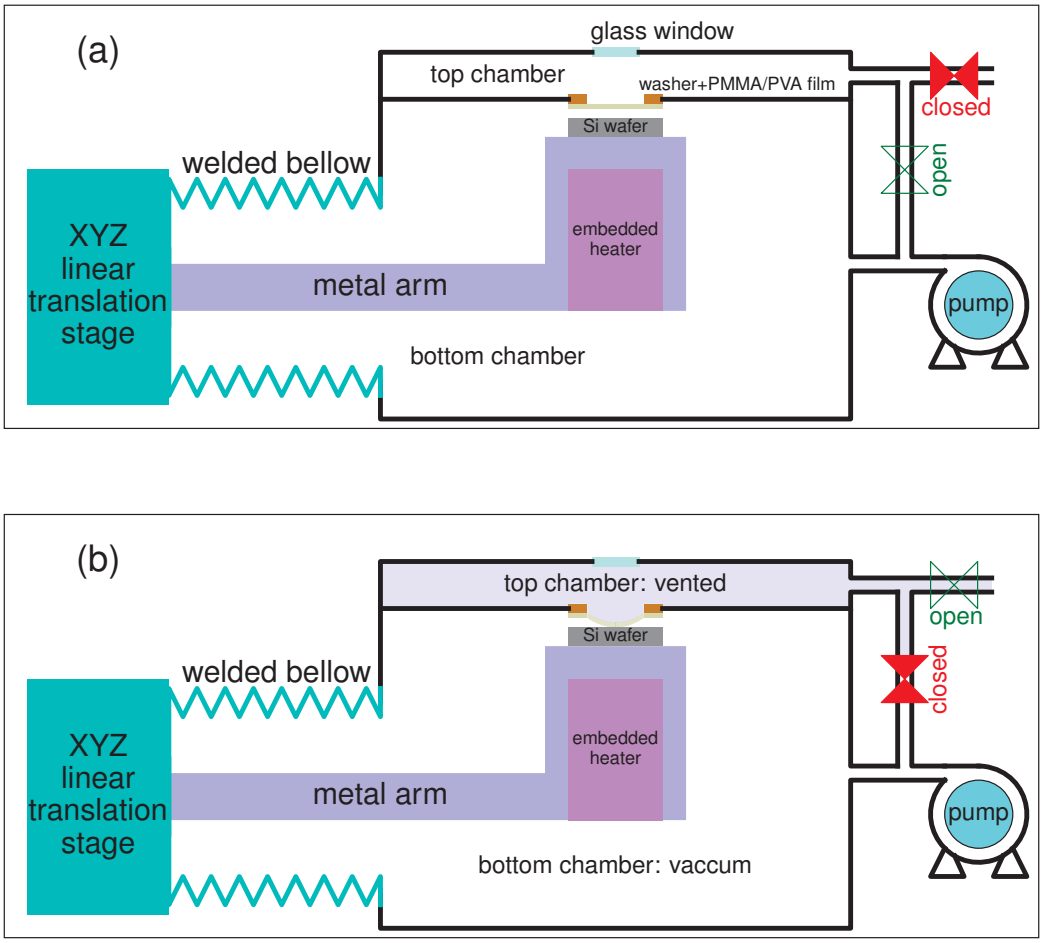

Fig. B.3. Operations of the flake transfer stage: (a) alignment and (b) chamber venting.

few mTorr. By adjusting the position of the silicon wafer using the XYZ translation stage, the flake on the PMMA/PVA film on the washer is aligned with the flake on the silicon wafer. The distance between them is roughly around 100 μm before the final contact. To make the final contact, as shown in Fig. B.3b, the top chamber is vented while the bottom chamber is still under vacuum. The pressure difference between two

chambers will push the PMMA/PVA film towards the silicon substrate so that firm contact is made. Heating the substrate to 120 ℃is suggested. The bottom chamber is then also vented, and the washer is detached from the metal separation. The silicon wafer with washer on top is placed into acetone bath to release the PVA film and washer. The flake transfer procedure is then completed. A gentle way to vent the top chamber is closing all the valves, so that only the bottom chamber is being pumped and the pressure difference is gradually built up.

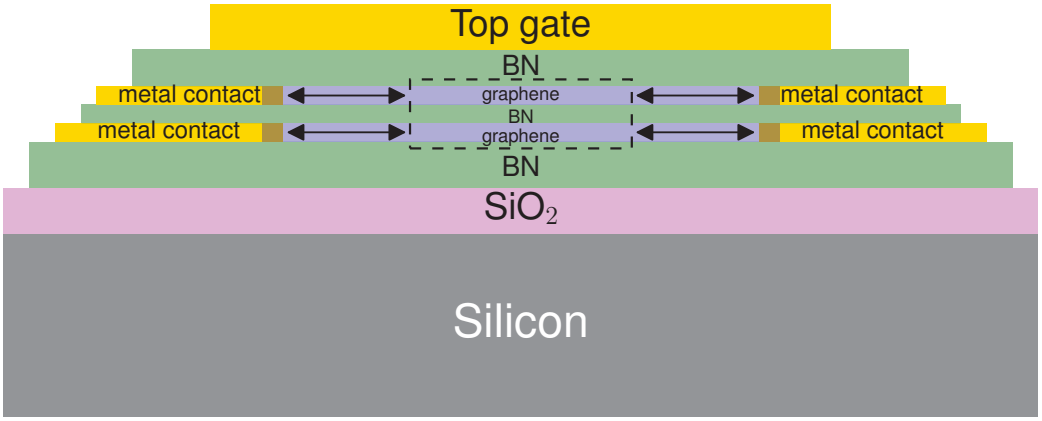

Fig. B.4. Cross section of the graphene double layer, viewed in the longitudinal direction of the Hall bar contained in the dashed box.

The major steps of the fabrication of a representative graphene double layer sample is shown in Fig. B.5, with every step explained as below:

(a) Graphene flake is exfoliated on PMMA/PVA film. Note that the parallel strips are on the opposite side of graphene.

(b) Bottom graphene flake is transferred onto the blueish boron nitride flake.

(c) Oxygen plasma etching defines the Hall bar of bottom graphene layer. Note that the boron nitride at the etched area become thinner, creating the shape of the Hall bar.

(d) Contacts to graphene and heater lines are made of Ti/Au.

(e) Boron nitride that will separate two graphene layers is exfoliated on PMMA/PVA film.

(f) Top graphene flake is exfoliated on PMMA/PVA film.

(g) The boron nitride of spacer in (e) and top graphene flake in (f) are transferred onto the Hall bar in (d).

(h) Top graphene flake is plasma etched into Hall bar and metal contacts and temperature sensors are made of Ti/Au.

(i) Boron nitride used for top gate dielectrics is exfoliated on PMMA/PVA film.

(j) The boron nitride flake in (i) is transferred on the device in (h) and the top gate is made.

(k) Zoom out (using 20x objective lens) of the device.

(l) Further zoom out (5x objective lens) of the device to include the contact pads.

The cross section of the device in Fig. B.5 is shown in Fig. B.4. In this cartoon the thickness of metal contacts is taken as the same as the thickness of graphene, which is not true in real samples where the thickness of the metal contact is usually about 50∼100 nm. If the metal contacts, e.g., to the bottom graphene layer, reach to the transverse edge of the Hall bar, the spacer boron nitride layer cannot uniformly touch the Hall bar area of the bottom graphene layer; there can exist a void gap between bottom graphene layer and the spacer boron nitride around the metal-graphene

contact area. This is obvious for the covering of thick boron nitride on metal contacts, i.e., shadows around metal contacts can be observed. To avoid this difficulty, metal contacts only touch the Hall bar leg that is made of graphene to ensure there is approximately 1 μm distance (indicated by the length of the double-head arrows in Fig. B.4) between metal contacts and graphene transverse edge.

We have used another graphene transfer recipe which is called the "pick-up" method. This method uses a small tip made of PDMS and PPC with a boron nitride flake on top. This boron nitride flake can either be picked-up by PPC or directly exfoliated onto PPC. This tip can then pick up the other flakes (e.g., boron nitride, graphene, or other 2D materials) prepared on clean SiO_2/Si chips. After building a stack of 2D materials, the contacts to graphene can be made through the exposed 1D graphene edge using plasma etching of the stack, a easy task for graphene single layer. However, if the stack contains the graphene double layer structure, it is difficult to make isolated contacts (particularly symmetrical Hall bar contacts for both graphene layers) to individual graphene layers. This procedure is very similar to fabricating the traditional double quantum well devices, where the quantum wells are grown first and the fabricating techniques are used to make the device. The trick of selective depletion allows isolated contacts to both layers, but it is not applicable to graphene since the conduction of graphene cannot be fully turned off using gate.

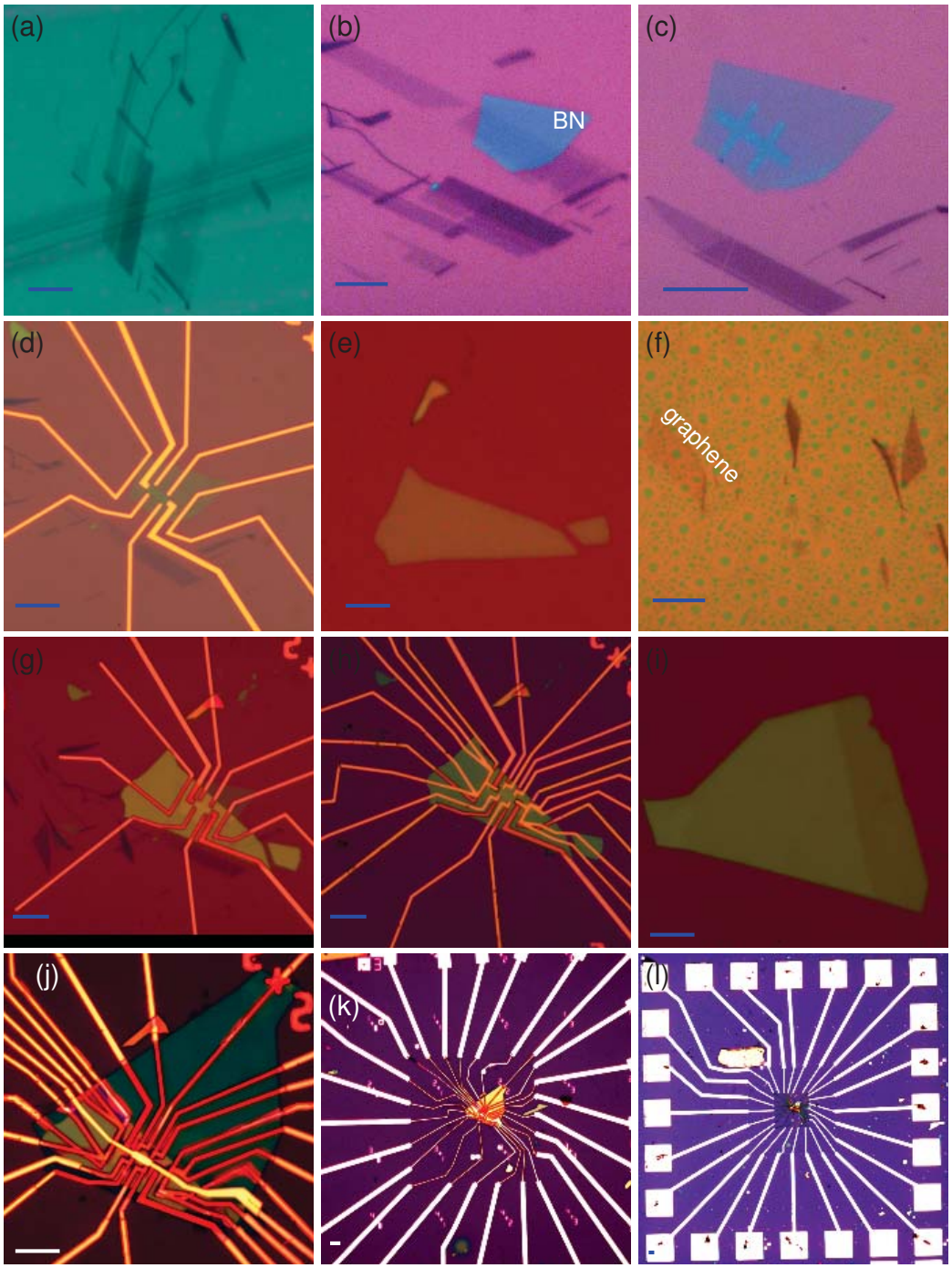

Fig. B.5. Fabrication procedure of graphene double layer. The scale bar in all subfigures is 10 μm.

C. CHARACTERIZATION OF GRAPHENE

Raman spectroscopy provides a convenient method to determine the quality (indicated by the intensity of D peak) and the layer number (indicated by the shape of the 2D peak, especially useful for single and bilayer graphene) of graphene flakes. In Fig. C.1 we show the Raman spectrum of single and bilayer graphene on boron nitride or SiO_2. An obvious feature is that the intensity ratio between the 2D and G peaks for graphene on boron nitride is larger than that for graphene on SiO_2. This is because using boron nitride as substrate can reduce the carrier doping in graphene and thus enhance the 2D peak intensity, [201] while the intensity of G peak is not sensitive to the level of carrier doping. The bilayer nature of graphene on boron nitride or SiO_2 is demonstrated by the decomposition of the 2D peak in (c) and (d) into the superposition of four Lorentzian peaks.

Two common kinds of metal contacts are fabricated to study the effect of thermal annealing on graphene-metal contacts, as shown in Fig. C.2. Obviously, thermal annealing dramatically increases the two-terminal resistance of Cr/Au contact (more than 10 times larger), while the two-terminal resistance of Ti/Au contact is increased about 2-3 times larger. We thus use Ti/Au as the metal contact to graphene for our samples.

Current annealing is a powerful yet dangerous tool to improve graphene-metal contact. The I-V curves for a representative two-terminal contact are shown in Fig. C.3. For the first annealing, the current slowly increases with two-terminal voltage initially. When the voltage is around 3.3 V, a sudden jump of the current shown in the

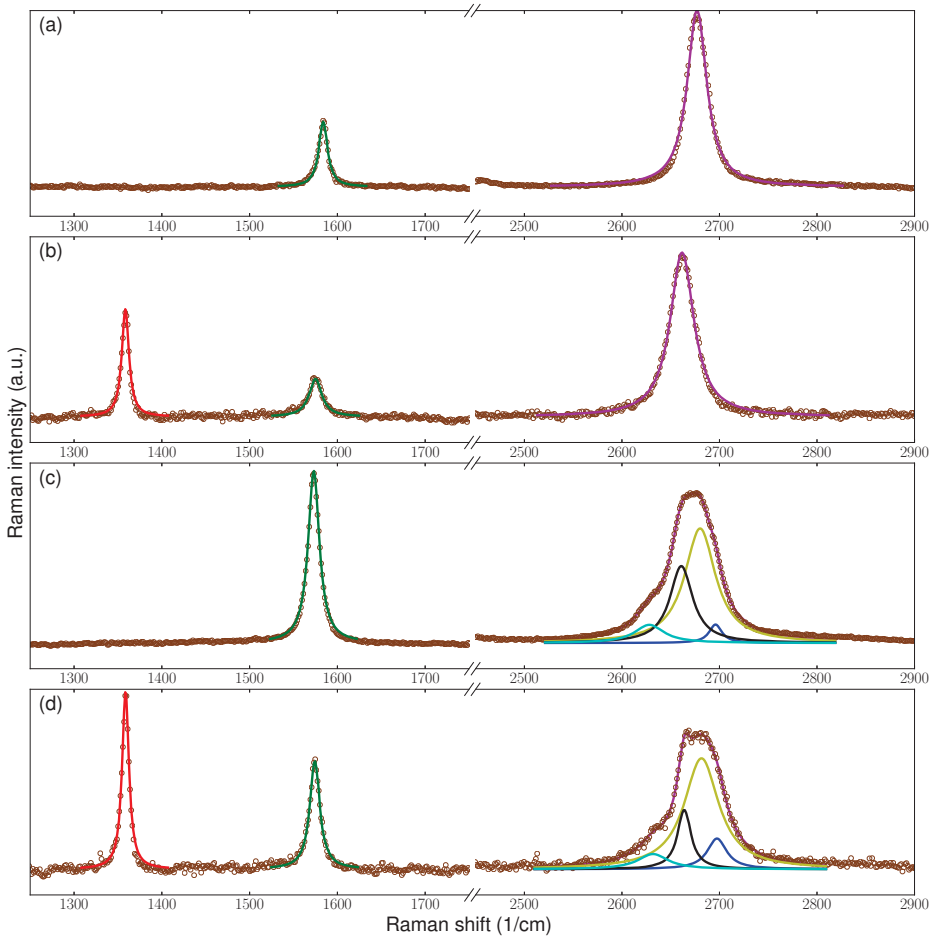

Fig. C.1. Raman spectrum of single layer graphene on SiO_2 (a) and boron nitride (b) and bilayer graphene on SiO_2 (c) and boron nitride (d).

dashed ellipse is observed, an indication of effective current annealing. On the second annealing, the low bias two-terminal resistance is dramatically reduced by more than one order, see Fig. C.3b.

The inset of Fig. C.4a presents a graphene double layer device without top gate. The curves in Fig. C.4a-c represent the field effect and quantum Hall measurement

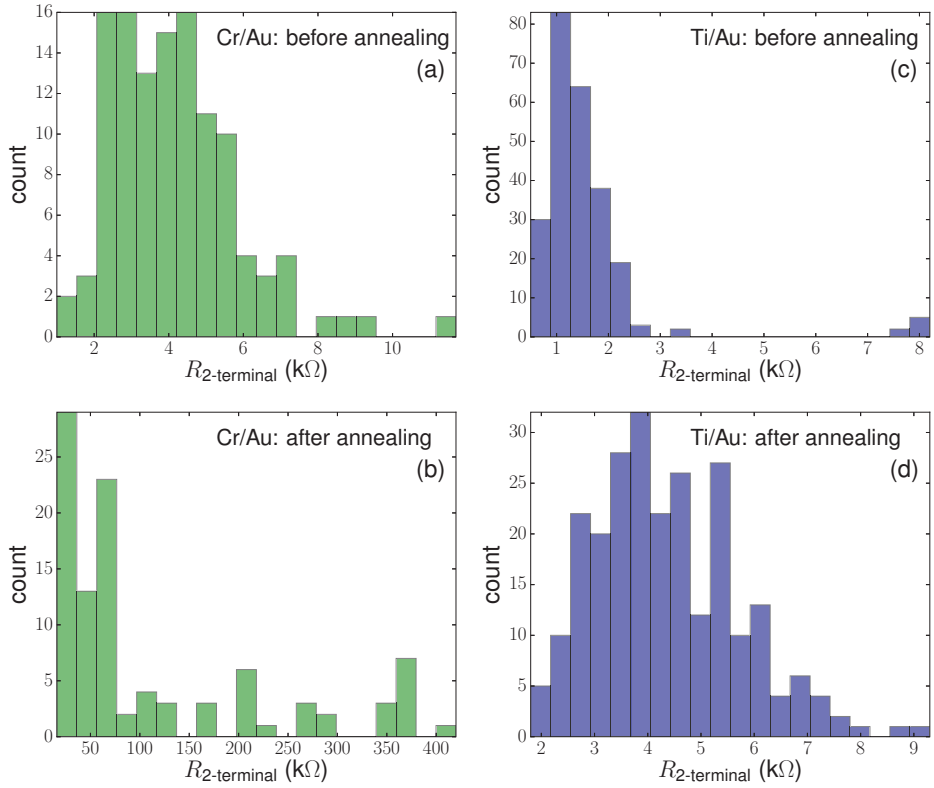

Fig. C.2. Histogram of two terminal resistance of graphene samples for contact metal of Cr/Au (a-b) and Ti/Au (c-d).

of the bottom layer graphene that is encapsulated between two boron nitride layers. The mobility of the bottom layer graphene at T = 300 mK is around 75000 cm^2/V-s, about one order of magnitude higher than that of the top layer graphene which is not covered by boron nitride and exposed. In Fig. C.4d, the field effect curve is shown for bilayer graphene sandwiched between two BN flakes fabricated using the pick-up method. Since the graphene surface is not exposed during sample fabrication, there is no hysteresis when the gate voltage is swept back and forth.

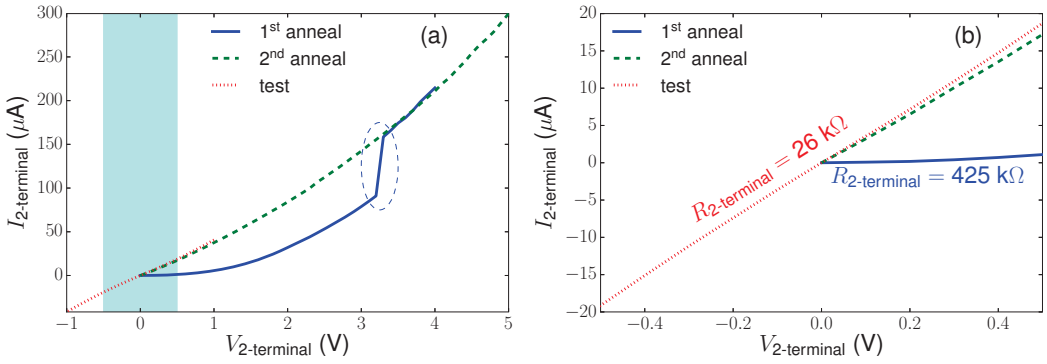

Fig. C.3. Current annealing (a) on a representative graphene sample and (b) is the zoom-in of the cyan region in (a).

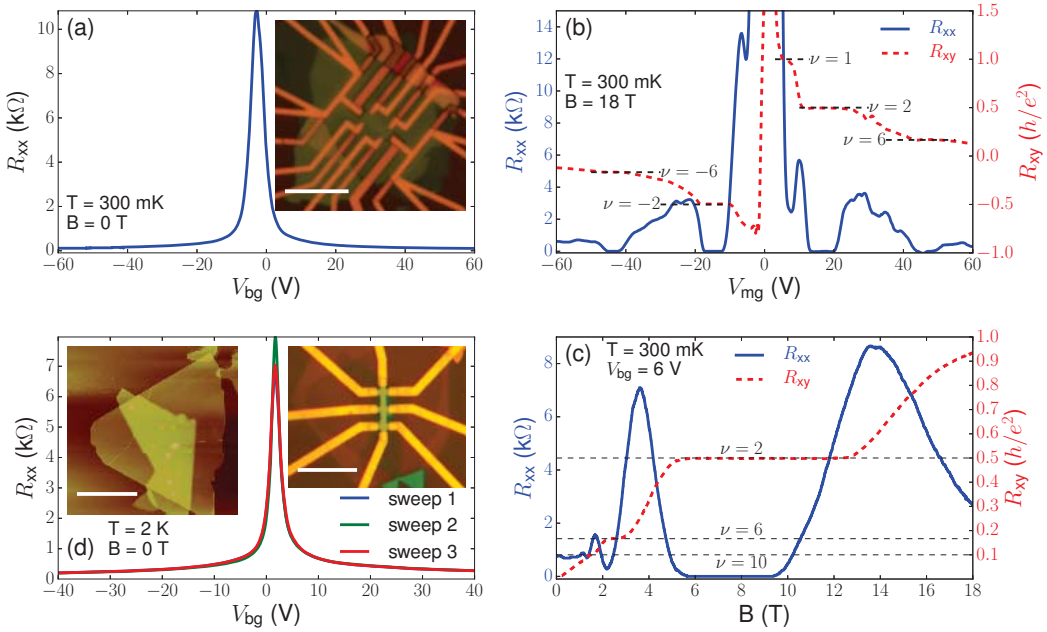

Fig. C.4. Field effect (a) and quantum Hall effect (b-c) of single layer graphene sandwiched between BN flakes and field effect (d) of bilayer graphene sandwiched between BN flakes. The scale bar in the insets of (a) and (d) is 10 μm.

D. TRANSPORT CHARACTERIZATIONS OF BILAYER GRAPHENE DOUBLE LAYER

D.1 Mobility measurements of individual layers

In order to characterize the quality of individual graphene layers, the magneto-transport at low magnetic field is presented in Fig. D.1 and Fig. D.2 for the bottom and top graphene layer respectively. Almost exact procedures in both figures are carried out. For example in Fig. D.1, the longitudinal resistivity $\rho_{\mathrm{b,xx}}$ and the transverse resistance $R_{\mathrm{b,xy}}$ vs. V_{bg} at $B = \pm 0.2$ T are plotted in (a). After shifting the charge neutral points to a common value, they are replotted in (b). The symmetrization for the longitudinal resistivity and the anti-symmetrization for the transverse resistance about magnetic field are performed and the resulting longitudinal resistivity and transverse resistance are shown in (c). All the subsequent analysis are based on the two curves in (c). In (d), the red circle (square) dots are for the electron (hole) density vs. V_{bg} calculated from the Hall effect formula $R_{\mathrm{b,xy}} = B/(n_{\mathrm{b}}e)$ for single type of carriers, and the gate capacitance and then the thickness of SiO_2 and boron nitride can be determined, shown at the bottom left in (d). The data near the charge neutral point is not used since the above formula is not valid. The blue solid curve in (d) is for the longitudinal conductivity. The field effect mobility μ_{FET} can be calculated by fitting the Drude formula $\sigma_{\mathrm{b,xx}} = (\mu_{\mathrm{FET}} n_{\mathrm{b}} e)^{-1}$, shown at the bottom right in (d). The carrier density dependent mobility is calculated using $\mu_{\mathrm{Hall,b}} = \sigma_{\mathrm{b,xx}}(n_{\mathrm{b}}e)^{-1}$ and is shown in (e). Another way to calculate mobility is by fitting only the longitudinal

resistivity to the formula $\rho_{\mathrm{b,xx}} = [e\mu_{\mathrm{FET}}\sqrt{[C_{\mathrm{bg}}(V_{\mathrm{bg}} - V_{\mathrm{D}})]^2 + n_0^2}]^{-1}$ where the mobility μ_{FET} is independent of the carrier density. Three different fitting results (fitting only electron side, only hole side and both sides simultaneously) are shown in (f). Note that if the same form of formula with different $\mu_{\mathrm{FET,e}}$ and $\mu_{\mathrm{FET,h}}$ is used to fit both electron and hole sides simultaneously, we must have $\mu_{\mathrm{FET,e}} = \mu_{\mathrm{FET,h}}$ if the fitted curve is required to be smooth at the charge neutral point. The key difference between the mobility calculation in (d) and (f) is that the charge carrier n_{b} used in (d) is from the Hall effect measurement while that in (f) is simply represented by $\sqrt{[C_{\mathrm{bg}}(V_{\mathrm{bg}} - V_{\mathrm{D}})]^2 + n_0^2}$. All three methods yield similar values of mobility at medium charge carrier density ($\sim 2 \times 10^{12}$ cm^{-2}).

In Fig. D.2, the transverse resistance in (c) for the top graphene layer exhibits large noise. It is thus filtered using the formula $R_{\mathrm{t,xy}} \sim 1/(V_{\mathrm{tg}} - V_{\mathrm{D}})$ and plotted as the dashed cyan line which is then used to calculate the density n_{t} in (d).

D.2 Consistency checks for Coulomb drag measurement

We first check the linear response of the Coulomb drag measurement, as shown in Fig. D.3. Both of the in-phase and out-of-phase components of the Coulomb drag voltage V_{d} are linearly dependent on the drive current I_{d} up to 3 μA. The low drive current will obviously lead to more noise, apparently shown in Fig. D.3d. We thus set the drive current in the following measurements as 2$\sim$3 μA.

We try a few different grounding configurations to study their effect particularly on the out-of-phase component of ρ_{d}. The contacts labels are shown in Fig. D.4a in order to clarify the grounding configurations. The in-phase and out-of-phase components of ρ_{d} are shown in Fig. D.4b and Fig. D.4c respectively. The in-phase component is not sensitive to the change of grounding contacts, however the out-of-phase com-

ponent strongly depends on the grounding configuration. If no more than one layer is grounded (blue and green curves in Fig. D.4c), the magnitude of the out-of-phase component is much smaller than that when both layers are grounded (red and cyan curves in Fig. D.4c), but it is almost identical to the mean Fig. D.4d of the red and cyan curves in Fig. D.4c. This is because the sign of the out-of-phase component due to capacitive coupling depends on the couping current direction which can be reversed by switching the grounding contact of the drag layer. In our following measurements, no more than one layer is grounded.

The in-phase and out-of-phase components of ρ_{d} for $V_{\mathrm{bg}} = 10$ V and $V_{\mathrm{bg}} = -10$ V when the drive and drag layers are swapped are shown in Fig. D.5. Both components except the out-of-phase component around the charge neutral point of top layer remain almost identical, consistent with the requirement by the Onsager's relation for the Coulomb drag resistivity. The slight difference infers that the conditions of the Onsager's relation are only approximately satisfied.

The lock-in frequency f dependence of ρ_{d} is shown in Fig. D.6a-b. Obviously the out-of-phase component strongly depends on f. This indicates that the appearance of the out-of-phase component could be due to some kind of capacitive coupling between two graphene layers. We calculate the ratio between the out-of-phase component and frequency, shown in Fig. D.6c. For f >7.7 Hz, the curves are almost sitting on each other, indicating that the magnitude of the out-of-phase component of ρ_{d} is proportional to f, thus the appearance of the out-of-phase component should be due to some kind of inductive coupling between two graphene layers. The sign of the out-of-phase component of ρ_{d} in Fig. D.3, Fig. D.6, Fig. D.4, Fig. D.5) is positive, which is consistent with the equivalent circuit to model the inductive coupling in Fig. D.6d. The mutual inductance is assumed to be M. We can assume an AC current $i_{\mathrm{d}} = I_0 \cos(2\pi f t)$ flowing in the top layer. The induced voltage due to the inductive

coupling on the bottom layer is then $v_{\rm d} = -M{\rm d}v_{\rm d}/{\rm d}t = 2\pi f M I_0 \sin(2\pi f t)$, and thus the out-of-phase component of $\rho_{\rm d}$ is $\rho_{\rm d,out\text{-}of\text{-}phase} = 2\pi f M$, proportional to the frequency f. Note that M is positive according to the theory of partial inductance. This model is not valid for very low frequency, e.g., the cyan curve in Fig. D.6c. Nonetheless, the out-of-phase component of $\rho_{\rm d}$ is dramatically suppressed by reducing f, e.g., at $f = 2.7$ Hz, the magnitude of the out-of-phase component is one order smaller than that of the in-phase component. We thus use low frequency (<5 Hz) drive current for all of the following Coulomb drag measurements.

D.3 Temperature sensor calibration

An AC current $I_{\rm h}$ of few mA and frequency ω is applied on the heater metal line. An small DC current $I_{\rm D}$ (0.5 mA in the following case) is applied on the sensor. The voltage probe on the sensor measures the voltage $V_{2\omega}$ at frequency 2ω. Suppose the four-terminal sensor resistance is $R_{\rm S}$ and its temperature coefficient is α. The temperature rise induced by $I_{\rm h}$ can be assumed to be δT at 2ω. Then the following relation holds:

$$V_{2\omega} = -I_{\rm D}\alpha\delta T R_{\rm S} = -I_{\rm D}\alpha a R_{\rm S} I_{\rm h}^2 \tag{D.1}$$

where δT is proportional to $I_{\rm h}^2$ and this proportional factor is denoted by a. Denote $A = I_{\rm D}\alpha a R_{\rm S}$ and it can be obtained by fitting the linear dependence of $V_{2\omega}$ on $I_{\rm h}^2$. Then we have $a = A(I_{\rm D}\alpha R_{\rm S})^{-1}$.

The Seebeck coefficient is then calculated from

$$S = \frac{V_{\rm S}}{g\Delta T} \tag{D.2}$$

where g is a geometric factor and V_{S} is the measured thermoelectric voltage between two ends of the sample, shown in Fig. D.7a. The temperature difference between two sensors created by the heater is

$$\Delta T = \frac{I_{\mathrm{h}}^2}{I_{\mathrm{D}}} \left(\frac{A}{\alpha R_{\mathrm{S}}}\bigg|_{\mathrm{c}} - \frac{A}{\alpha R_{\mathrm{S}}}\bigg|_{\mathrm{f}} \right). \tag{D.3}$$

The subscripts $_{\mathrm{c}}$ and $_{\mathrm{f}}$ are for the quantities (A, α and R_{S}) measured for the closer and further temperature sensors from the heater respectively. For example for the heater 4-13 of the device in Fig. D.7a, $_{\mathrm{c}}$ is for the sensor 3-2-16-12 and $_{\mathrm{f}}$ is for the sensor 6-8-9-10.

The uncertainty δS of S is

$$\delta S = \frac{V_{\mathrm{s}}}{g\Delta T^2} \frac{I_{\mathrm{h}}^2}{I_{\mathrm{D}}} \sqrt{\left(\frac{\delta A^2}{\alpha^2 R_{\mathrm{S}}^2} + \frac{\delta \alpha^2 A^2}{\alpha^4 R_{\mathrm{S}}^2} + \frac{\delta R_{\mathrm{S}}^2 A^2}{\alpha^2 R_{\mathrm{S}}^4} \right)\bigg|_{\mathrm{c}} + \left(\frac{\delta A^2}{\alpha^2 R_{\mathrm{S}}^2} + \frac{\delta \alpha^2 A^2}{\alpha^4 R_{\mathrm{S}}^2} + \frac{\delta R_{\mathrm{S}}^2 A^2}{\alpha^2 R_{\mathrm{S}}^4} \right)\bigg|_{\mathrm{f}}} \tag{D.4}$$

where δA, $\delta\alpha$ and δR_{S} are the measured uncertainties of A, α and R_{S} respectively.

For the device shown in Fig. 4.1, the length of both heaters is much larger than the width or length of graphene channel, thus we can assume one dimensional temperature profile created by the heater. We adopt such design for all of our samples in order to simplify the data analysis. The temperature sensor is located further from graphene channel than the heater, because this is convenient for the sample fabrication. We also adopt such design that the distance between the right (left) heater and right (left) edge of graphene channel is the same as that between the right (left) heater and right (left) temperature sensor for most of our devices. Thus the right (left) sensor effectively measures the temperature rise at the right (left) edge of graphene channel when the heating current is applied to the right (left) heater. Here we assume the existence of the graphene and boron nitride flakes and metal lines has negligible effect on the temperature profile, i.e., the temperature profile is primarily determined by

the properties of the substrate. This assumption can be justified since the substrate is bulky and has small thermal resistance while the graphene and boron nitride flakes are small and thin and have large thermal resistance.

The geometrical factor in Eq. (D.2) can be determined by the parameters L and L_{S} in Fig. D.7a. For small temperature rises (usually < 1 K), we can reasonably assume that $g = L_{\mathrm{S}}/L$.

The temperature coefficient α of the sensors can be calculated from the temperature dependence of the sensor resistance R_{s} in Fig. D.7b. The value of α is calculated by doing a linear fitting in the ranges represented by the red areas in Fig. D.7b while R_{S} is taken as the value of the fitted line at the required temperature. The linear fitting can also provide the uncertainty of α. The uncertainty of R_{S} is equal to the square root of the variance of the difference between the measurement and fitting results. In Fig. D.7c and Fig. D.7d we plot the curves of $V_{2\omega}$ vs. I_{h}^2 for the closer and further sensors respectively. The out-of-phase component of $V_{2\omega}$ and its I_{h}^2 dependence are negligibly small, as expected. The coefficient A and its uncertainty are calculated by the linear fitting and the fitted result is plotted as the dashed curve.

The measured A, α and R_{S} and their uncertainties δA, $\delta\alpha$ and δR_{S} for the device in Fig. 4.1 are presented in Table (D.1). In the last row, the temperature difference ΔT between two sensors increases when the temperature is reduced. This is because the heater current is $I_{\mathrm{h}} \sim 4$ mA, approximately same for the four temperatures, leading to approximately same heating power for them. On the other hand, the thermal conductivity of the substrate decreases with temperature. Therefore larger ΔT is created at lower temperature.

Table D.1.
Measured parameters to calculate Seebeck coefficient for the device in Fig. 4.1. The top (bottom) line for each row is for the closer (further) sensor, except the last row of the temperature difference between two sensors.

T (K)	50	100	200	296
$\alpha \pm \delta\alpha$	7.476±0.068	3.478±0.062	1.181±0.027	1.223±0.006
($10^{-4}\mathrm{K}^{-1}$)	7.577±0.067	3.505±0.065	1.053±0.025	1.045±0.006
$R_\mathrm{S} \pm \delta R_\mathrm{S}$	97.005±0.045	99.617±0.046	101.535±0.048	102.601±0.046
(Ω)	99.173±0.046	101.786±0.050	103.641±0.046	104.620±0.047
$A \pm \delta A$	8.0457±0.0426	2.2003±0.0063	0.5790±0.0004	0.6344±0.0004
($\mu\mathrm{V/mA}^2$)	0.7787±0.0014	0.2553±0.0009	0.1023±0.0003	0.1469±0.0003
ΔT (K)	3.84	2.05	1.31	1.15

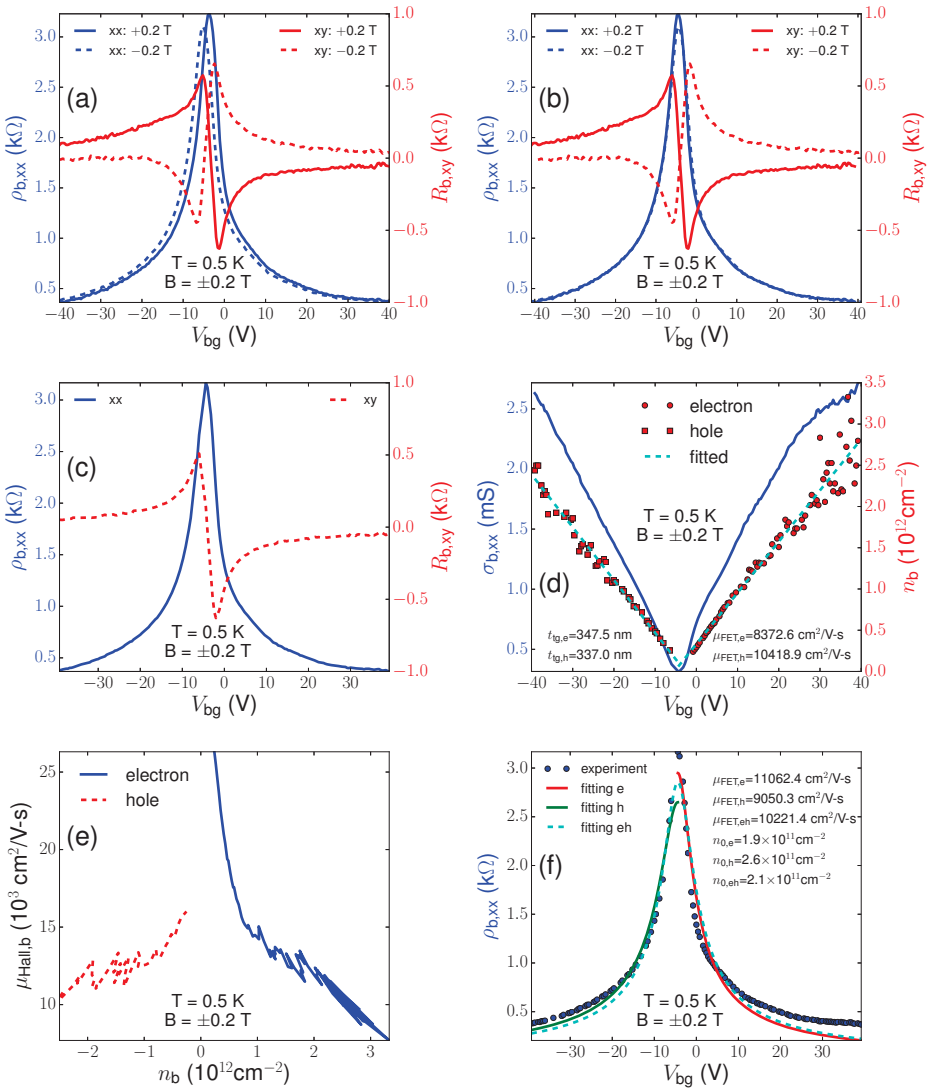

Fig. D.1. Longitudinal resistivity $\rho_{\mathrm{b,xx}}$ and Hall resistance $R_{\mathrm{b,xy}}$ (a) of bottom BLG layer vs. V_{bg} measured at B = +0.2 T and B = −0.2 T, (b) $\rho_{\mathrm{b,xx}}$ and $R_{\mathrm{b,xy}}$ vs. V_{bg} at B = +0.2 T and B = −0.2 T with their charge neutral points moved to the mean of that in (a), (c) symmetrized $\rho_{\mathrm{b,xx}}$ (= ($\rho_{\mathrm{b,xx}}$(B=+0.2 T) + $\rho_{\mathrm{b,xx}}$(B=-0.2 T))/2) and anti-symmetrized $R_{\mathrm{b,xy}}$ (= ($R_{\mathrm{b,xy}}$(B=+0.2 T) − $R_{\mathrm{b,xy}}$(B=-0.2 T))/2) vs. V_{bg}, (d) longitudinal conductivity $\sigma_{\mathrm{b,xx}}$ and carrier density n_{b} vs. V_{bg}, (e) Hall mobility $\mu_{\mathrm{Hall,b}}$ vs. n_{b}, and (f) fitting formula $\rho_{\mathrm{b,xx}} = [e\mu_{\mathrm{FET}}\sqrt{[C_{\mathrm{bg}}(V_{\mathrm{bg}} - V_{\mathrm{D}})]^2 + n_0^2}]^{-1}$ to measured $\rho_{\mathrm{b,xx}}$ vs. V_{bg} in (c).

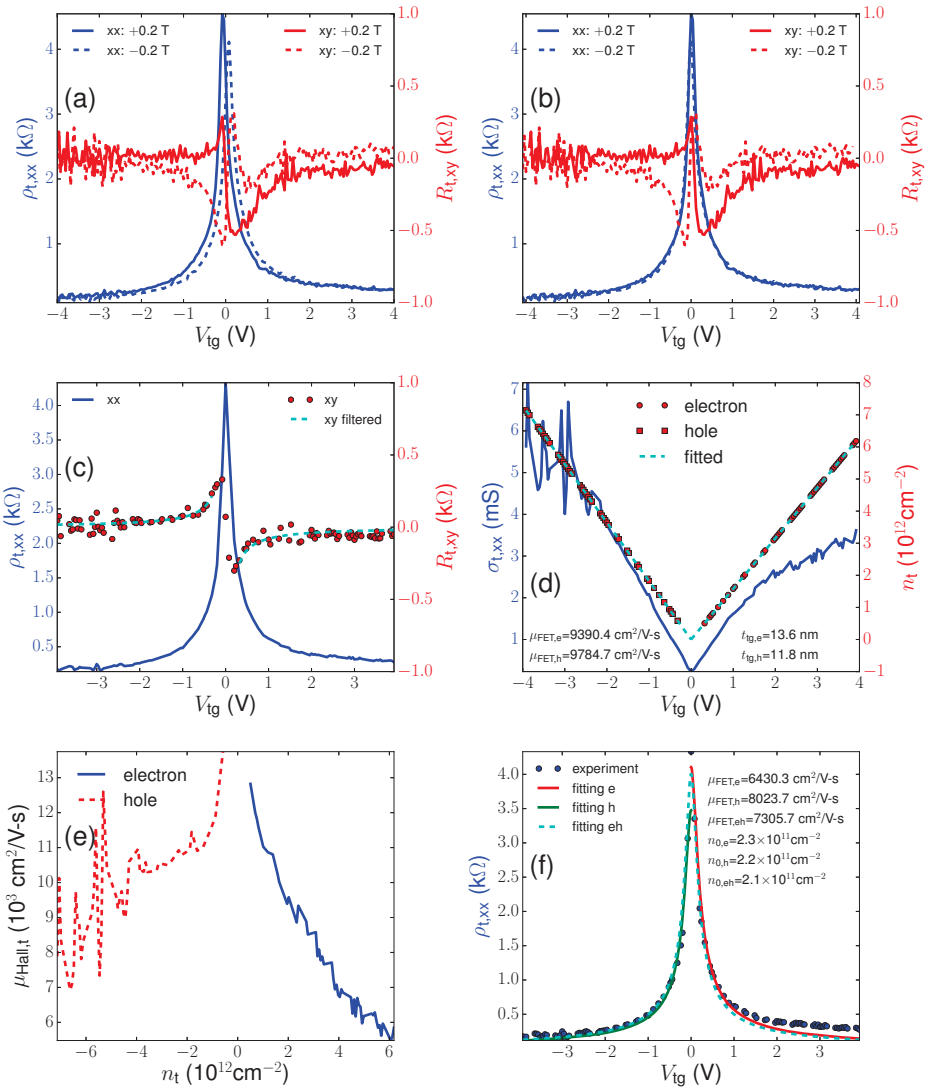

Fig. D.2. Longitudinal resistivity $\rho_{t,xx}$ and Hall resistance $R_{t,xy}$ (a) of top BLG layer vs. V_{tg} measured at B = +0.2 T and B = −0.2 T, (b) $\rho_{t,xx}$ and $R_{t,xy}$ vs. V_{tg} at B = +0.2 T and B = −0.2 T with their charge neutral points moved to the mean of that in (a), (c) symmetrized $\rho_{t,xx}$ (= ($\rho_{t,xx}$(B=+0.2 T)+$\rho_{t,xx}$(B=-0.2 T))/2) and anti-symmetrized $R_{t,xy}$ (= ($R_{t,xy}$(B=+0.2 T) − $R_{t,xy}$(B=-0.2 T))/2) vs. V_{tg}, (d) longitudinal conductivity $\sigma_{t,xx}$ and carrier density n_t vs. V_{tg}, (e) Hall mobility $\mu_{Hall,t}$ vs. n_t, and (f) fitting formula $\rho_{t,xx} = [e\mu_{FET}\sqrt{[C_{tg}(V_{tg} - V_D)]^2 + n_0^2}]^{-1}$ to measured $\rho_{t,xx}$ vs. V_{tg} in (c). The cyan dashed line in (c) is the fitted result of formula $R_{t,xy} \sim 1/(V_{tg} - V_D)$.

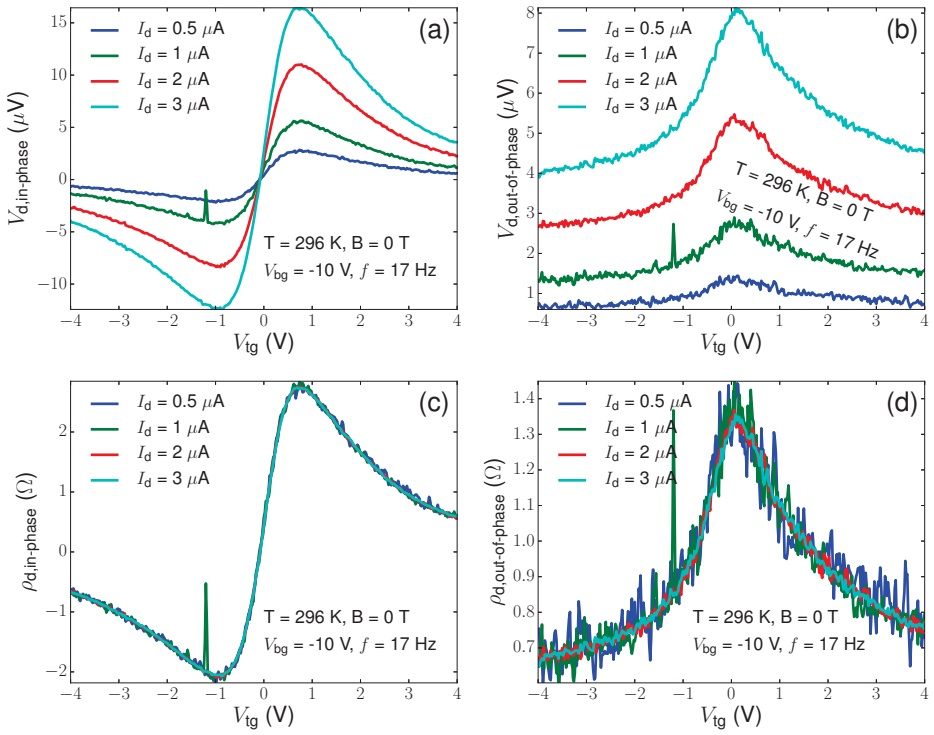

Fig. D.3. In-phase (a) and out-of-phase (b) components of Coulomb drag voltage V_d measured on top graphene layer while the drag current I_d is driven on the bottom graphene layer and in-phase (c) and out-of-phase (d) components of Coulomb drag resistivity $\rho_\mathrm{d} = V_\mathrm{d}/I_\mathrm{d}$.

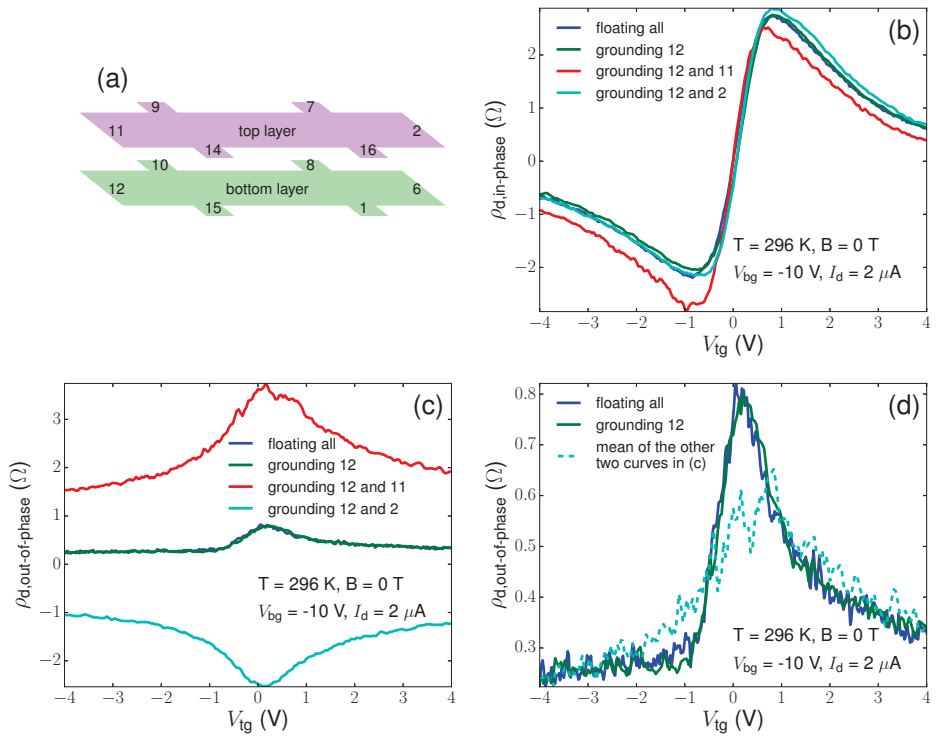

Fig. D.4. Contact labels (a) for both graphene layers, in-phase (b) and out-of-phase (c) components of Coulomb drag resistivity ρ_d measured at different grounding conditions and (d) blue and green curve and the mean (dashed cyan) of red and cyan curves in (c).

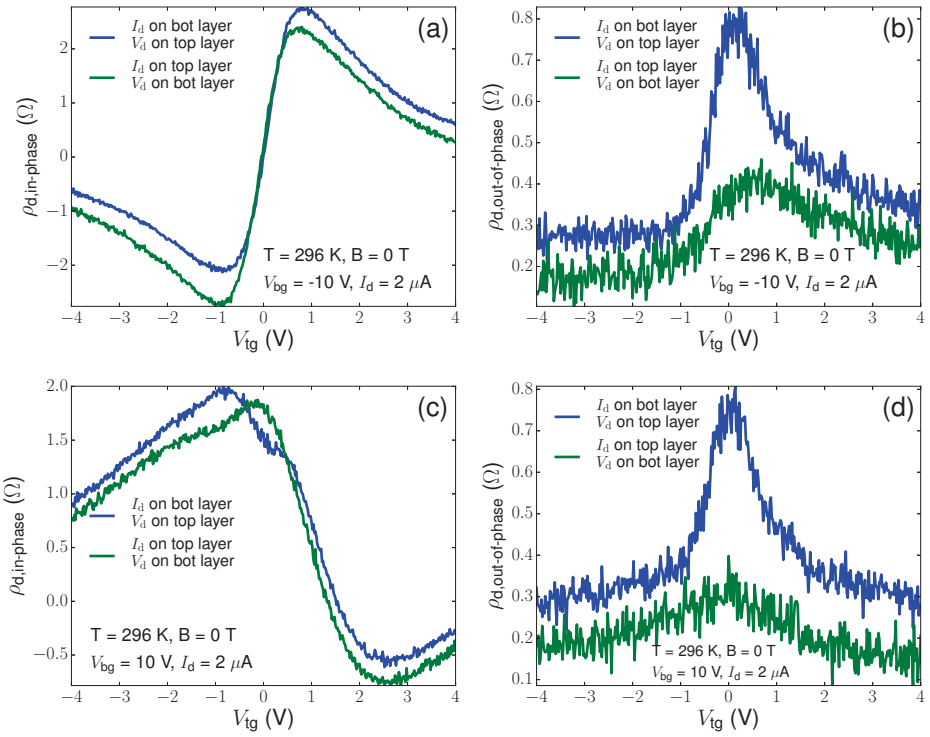

Fig. D.5. In-phase (a) and out-of-phase (b) components of Coulomb drag resistivity ρ_d at V_{bg} = -10 V and in-phase (c) and out-of-phase (d) components of ρ_d at V_{bg} = 10 V, when the drive and drag layers are swapped.

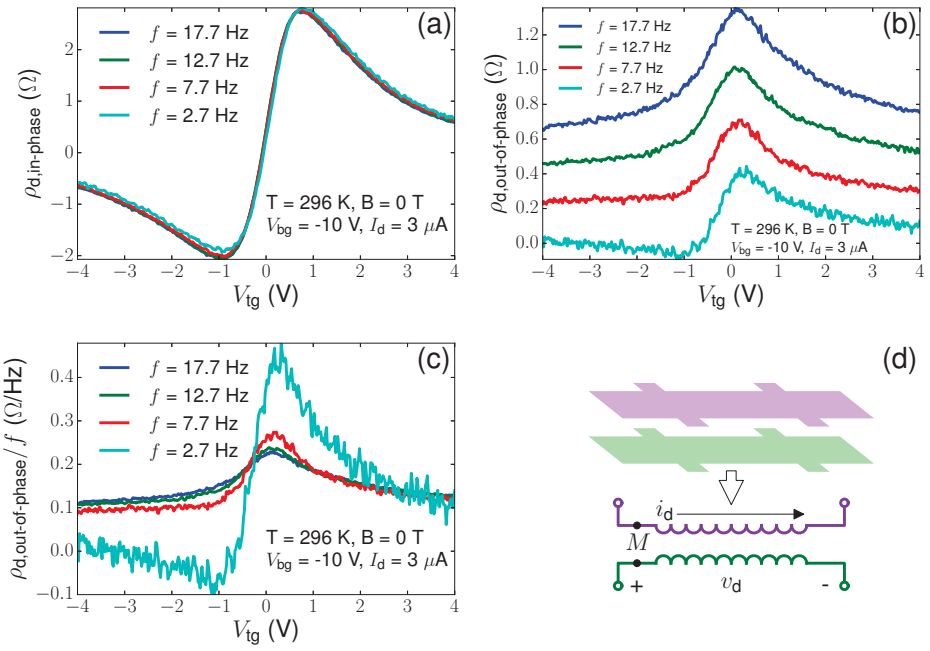

Fig. D.6. In-phase (a) and out-of-phase (b) components of Coulomb drag resistivity ρ_{d} measured at different lock-in frequencies, (c) the out-of-phase component of ρ_{d} normalized by the lock-in frequency and (d) the equivalent circuit to model the inductive coupling between two layers.

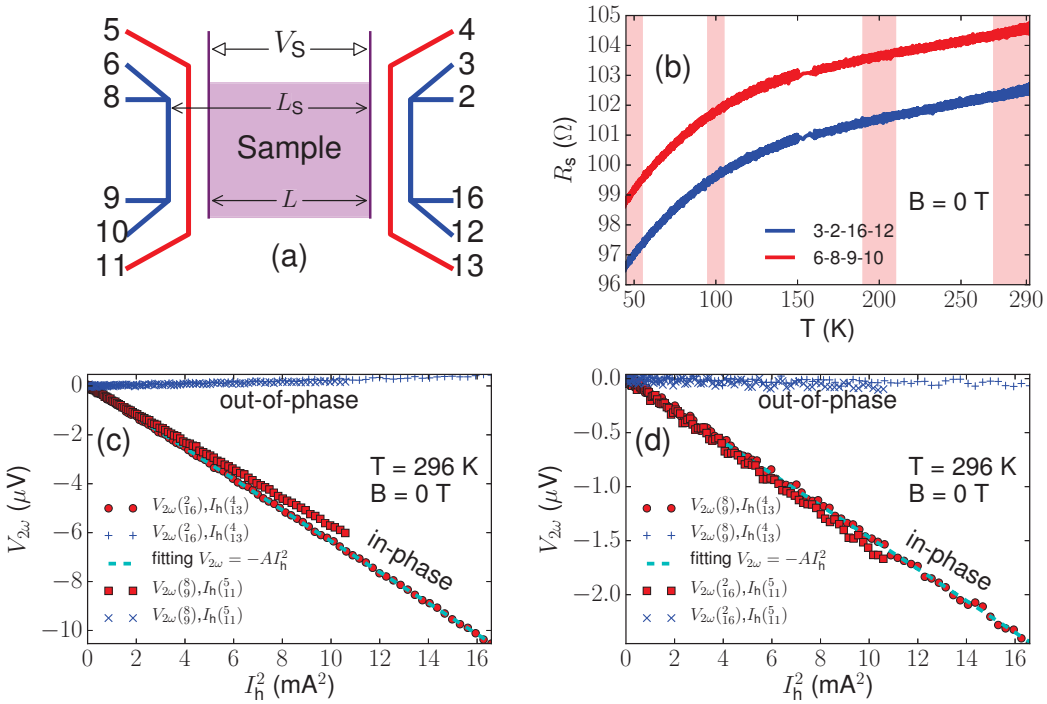

Fig. D.7. Configuration (a) of sample (purple area), heaters (red lines) and temperature sensors (blue lines), temperature dependent four-terminal resistance R_s of the temperature sensors in (a), 2ω voltage (c) between the two inner leads (8-9 or 2-16) of the sensor with heating current applied on the nearby (5-11 or 4-13) heater, and 2ω voltage (c) between the two inner leads (8-9 or 2-16) of the sensor with heating current applied on the further (4-13 or 5-11) heater.

E. TRANSFORMER-BASED SETUP FOR DC-ISOLATED AC CURRENT SOURCE

In some circumstances a transformer has to be used to supply DC-isolated AC current. For example, for double layer system with only one (e.g., back) gate available, in order to tune the carrier densities of the two layers simultaneously, except the gate voltage applied between the gate and the layer closer to it, another mutual gate voltage has also to be applied between two layers. If now both layers are connected to some instruments (e.g., two lock-ins to measure the intralayer resistivities or Coulomb drag measurement), the mutual gate voltage will create a current path through those instruments and thus it become incapable of tunning the carrier densities.

In Coulomb drag measurement, the setup in Fig. E.1[1] can supply AC current on the sample connected on the out branch of the transformer (green area) at S+ and S- with a resistance R_{S}. The current can be controlled by voltage between IN+ and IN- on the in branch. The resistances R_{O} and R_{I} are used to measure the current flowing in the out and in branches respectively.

To characterize this setup, first of all, an AC voltage of $V_{\mathrm{oscv}} = 1$ V is applied to IN+ and IN-. The voltages V_{TI} and V_{SI} are measured at TI+/- and SI+/- respectively, and they are plotted in the first column in Fig. E.2 for different value of R_{S}, vs. the frequency f_{oscv} of V_{oscv}. Around $f_{\mathrm{oscv}} = 40$ Hz, the out-of-phase components of V_{TI} and V_{SI} are close to zero. In the second column, $V_{\mathrm{SI}}/V_{\mathrm{TI}}$ is calculated, and the magnitude and in-phase ratios become same for $f_{\mathrm{oscv}} \sim 40$ Hz. We then fix the frequency at

[1]The transformer here is from jensen: http://www.jensen-transformers.com/product/sub-1rr/

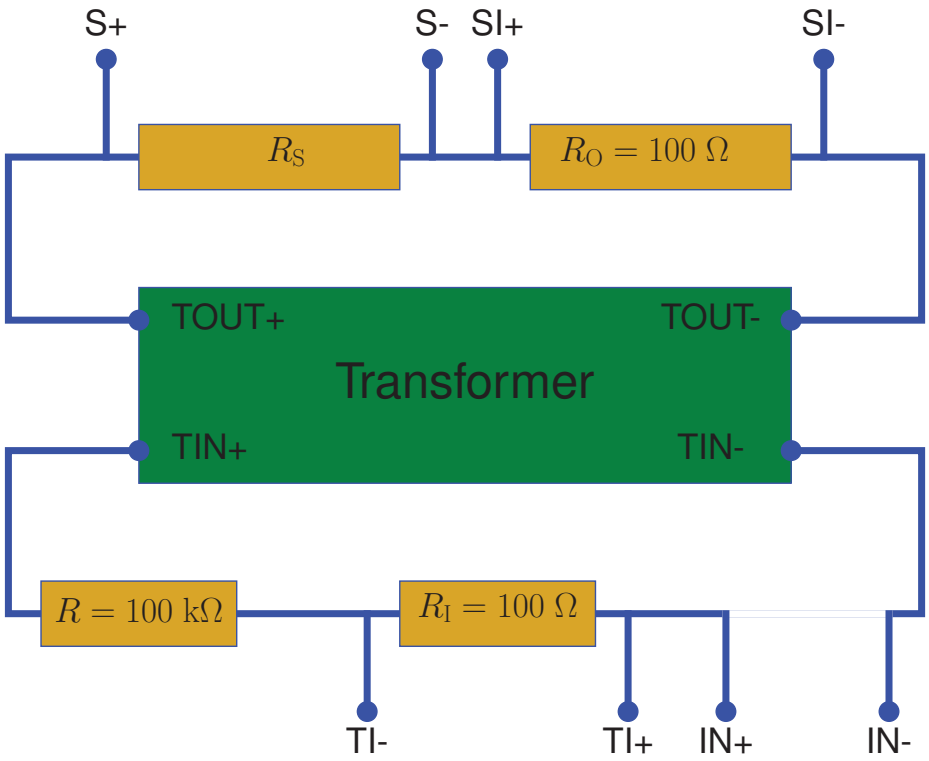

Fig. E.1. Measurement setup based on transformer.

$f_{\mathrm{oscv}} = 40$ Hz, and measure the ratio $V_{\mathrm{SI}}/V_{\mathrm{TI}}$ vs. V_{oscv}, shown in the third column. Obviously this ratio is independent of the value of the voltage source V_{oscv}.

The ratio $V_{\mathrm{SI}}/V_{\mathrm{TI}}$ vs. R_{S} at $f_{\mathrm{oscv}} = 40$ Hz is plotted in Fig. E.3. In order to know the current flowing in the out branch of the transformer, the two-terminal resistance R_{S} can be measured first to determine the value of $V_{\mathrm{SI}}/V_{\mathrm{TI}}$. Then the voltage V_{TI} can be measured to calculated the current across the sample.

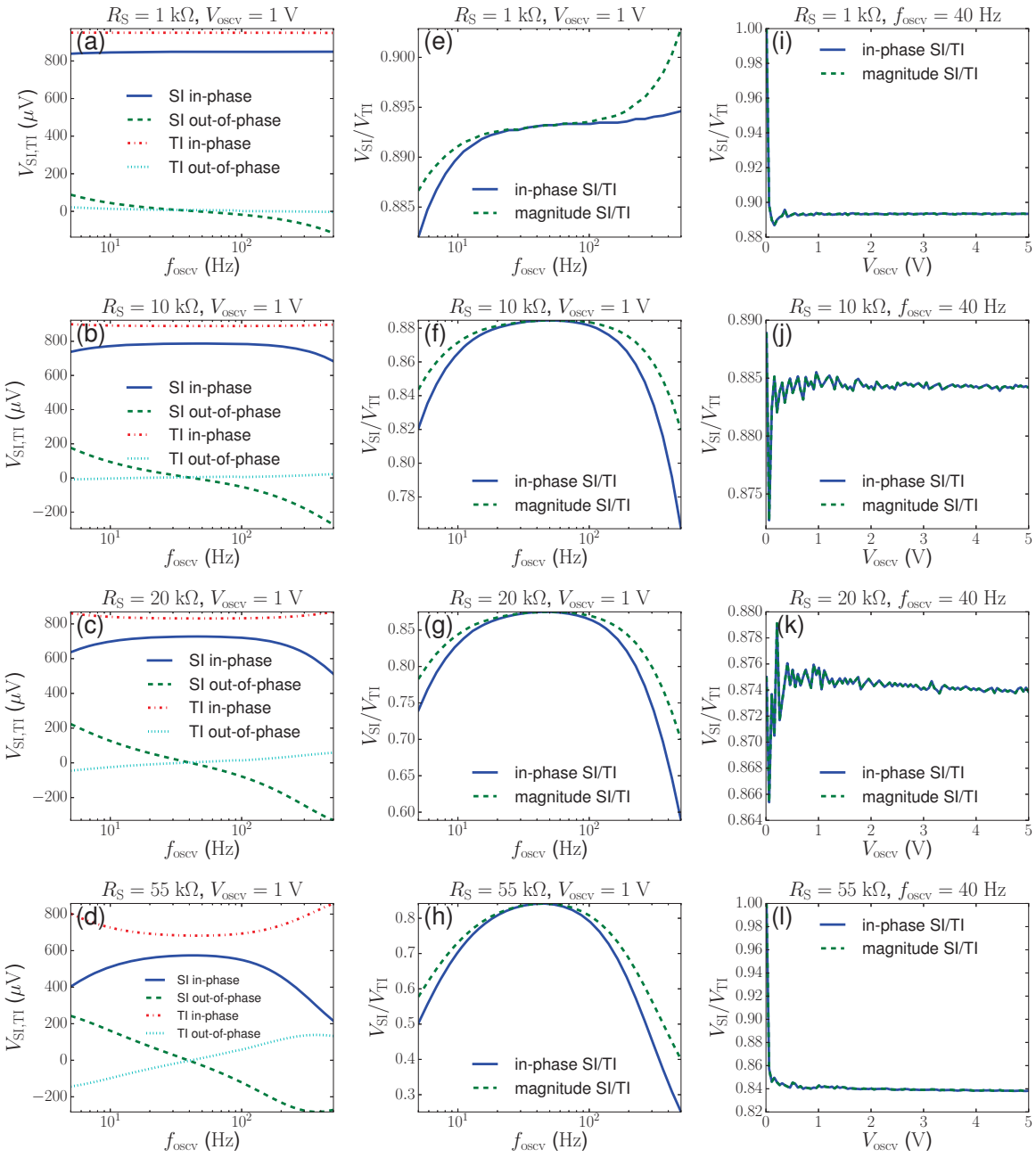

Fig. E.2. Voltages V_{SI} and V_{TI} vs. f_{oscv} (a-d), the ratio V_{SI}/V_{TI} vs. f_{oscv} (e-h) and V_{SI}/V_{TI} vs. V_{oscv} (i-l) for four different values of R_S labeled in the title of each subplot.

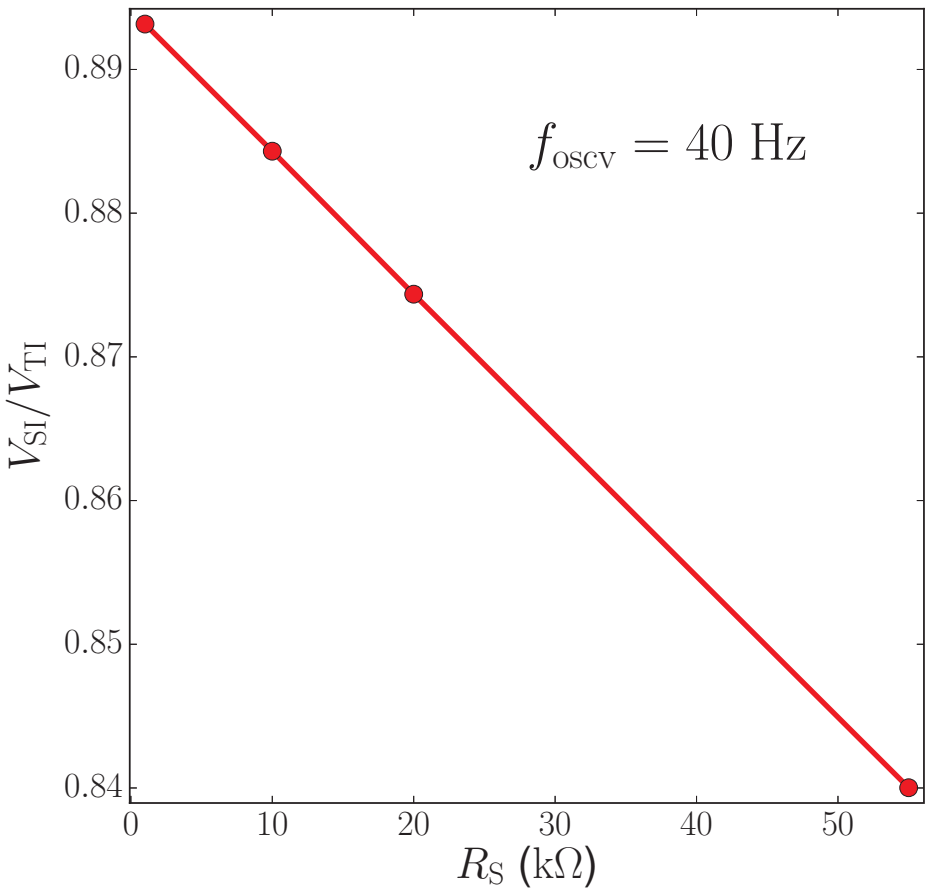

Fig. E.3. Ratio $V_{\mathrm{SI}}/V_{\mathrm{TI}}$ vs. R_{S} at $f_{\mathrm{oscv}} = 40$ Hz.

F. GRAPHENE/BN/NIOBIUM DISELENIDE

In Fig. F.1, the top view (without top gate BN and metal) and cross section view of graphene/BN/$NbSe_2$ samples are shown in (a) and (b) and the optical image of a representative sample is shown in (c).

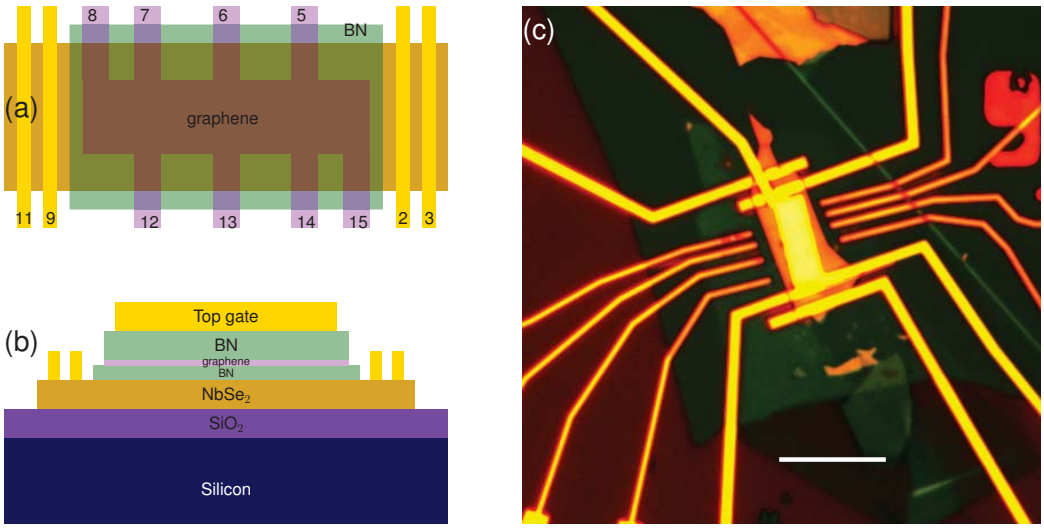

Fig. F.1. Top (a) and cross section (b) views of graphene/BN/$NbSe_2$ samples and the optical image (c) of the measured sample. The scale bar of white line segment in (c) is 10 μm.

For the sample shown in Fig. F.1c, the thickness of the BN between graphene and $NbSe_2$ is 5 nm. The four-terminal resistance of graphene are shown in Fig. F.2a for three different temperatures at zero magnetic field and the quantum Hall effect at T = 1.7 K and B = 6 T is shown in Fig. F.2b. The single layer nature of graphene can

be clearly identified from the Hall plateaus. In Fig. F.2c-d, the T and B dependent resistances of the $NbSe_2$ flake are shown and we can clearly determine the superconducting transition temperature T_c and the critical magnetic field B_c as $\sim$ 7 K and $\sim$ 3.3 T respectively.

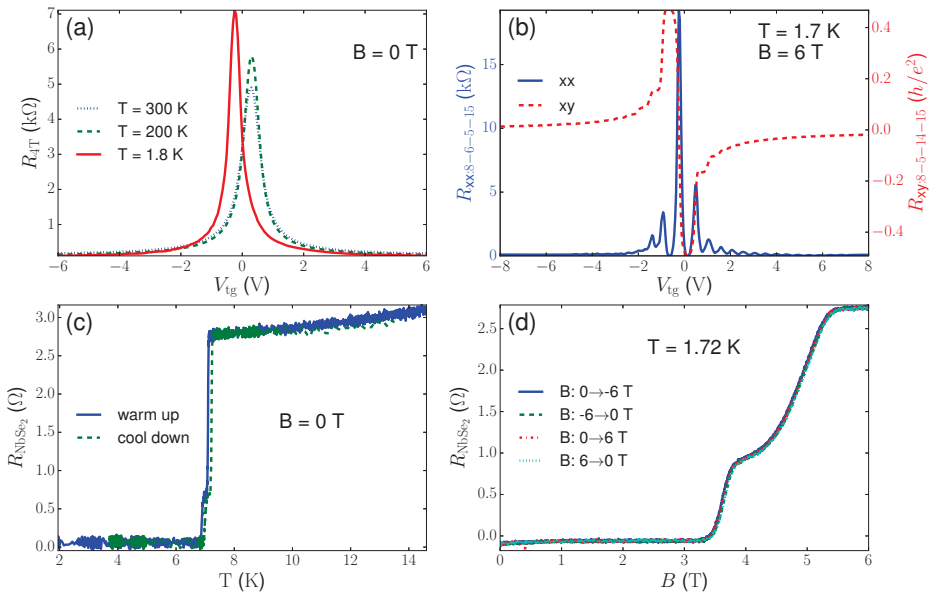

Fig. F.2. Field effect (a) and quantum Hall effect (b) of graphene and T (c) and B (d) dependence of the resistance of $NbSe_2$ flake.

At two different temperatures of 15 K (a-c) and 1.7 K (d-f), the color maps of the graphene longitudinal resistance (a,d), graphene transverse resistance (b,e) and the resistance of $NbSe_2$ (c,f) vs. top gate voltage V_{tg} and magnetic field B are shown in Fig. F.3.

In Fig. F.4, at two different temperatures of 8 K (a-c) and 6 K (d-f), the in-phase (a,d) and out-of-phase (b,e) components of the Coulomb drag voltage between graphene and $NbSe_2$ vs. V_{oscv} (controlling the drag current) and the drag resistance

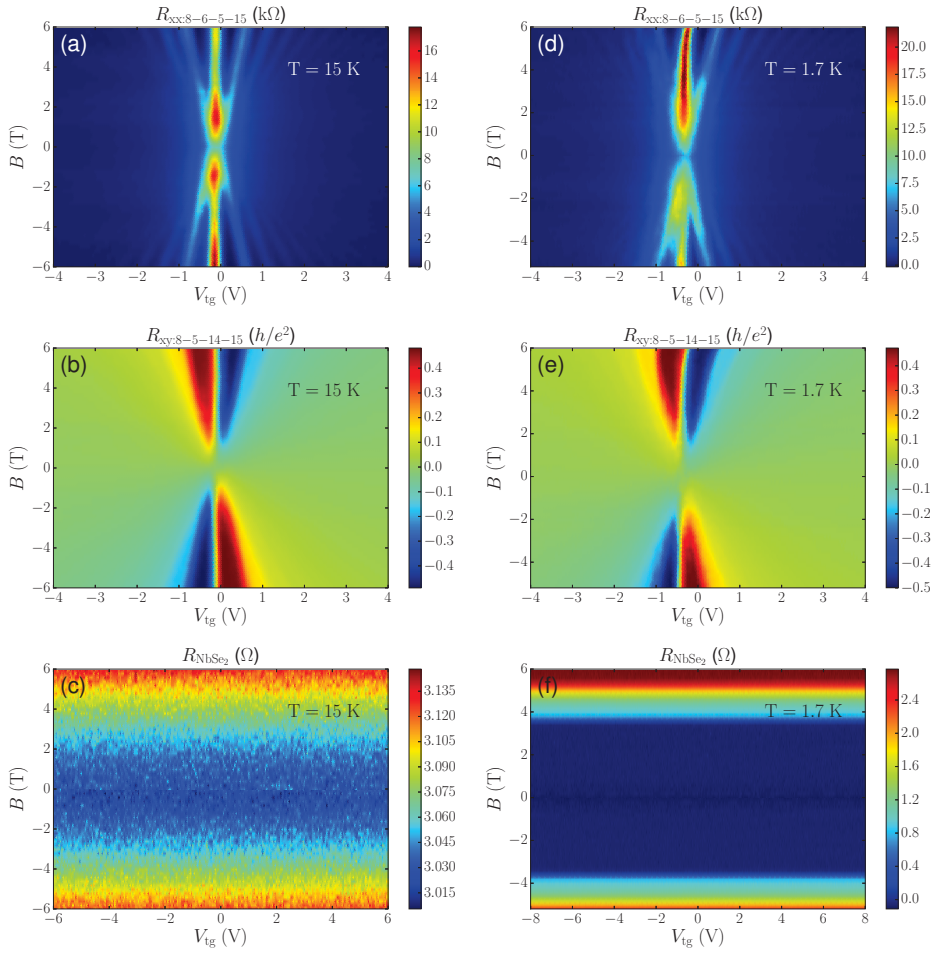

Fig. F.3. Color maps of graphene longitudinal resistance (a,d), graphene transverse resistance (b,e) and the resistance of $NbSe_2$ (c,f) vs. V_{tg} and B at T = 15 K (a-c) and T = 1.7 K (d-f).

(c,f) vs. V_{tg} are shown. Obviously, the Coulomb drag voltage is linearly dependent on the AC voltage supply V_{oscv} and thus the drag resistance can be easily calculated. Note that the magnitude of the out-of-phase component is much smaller than that of the in-phase component.

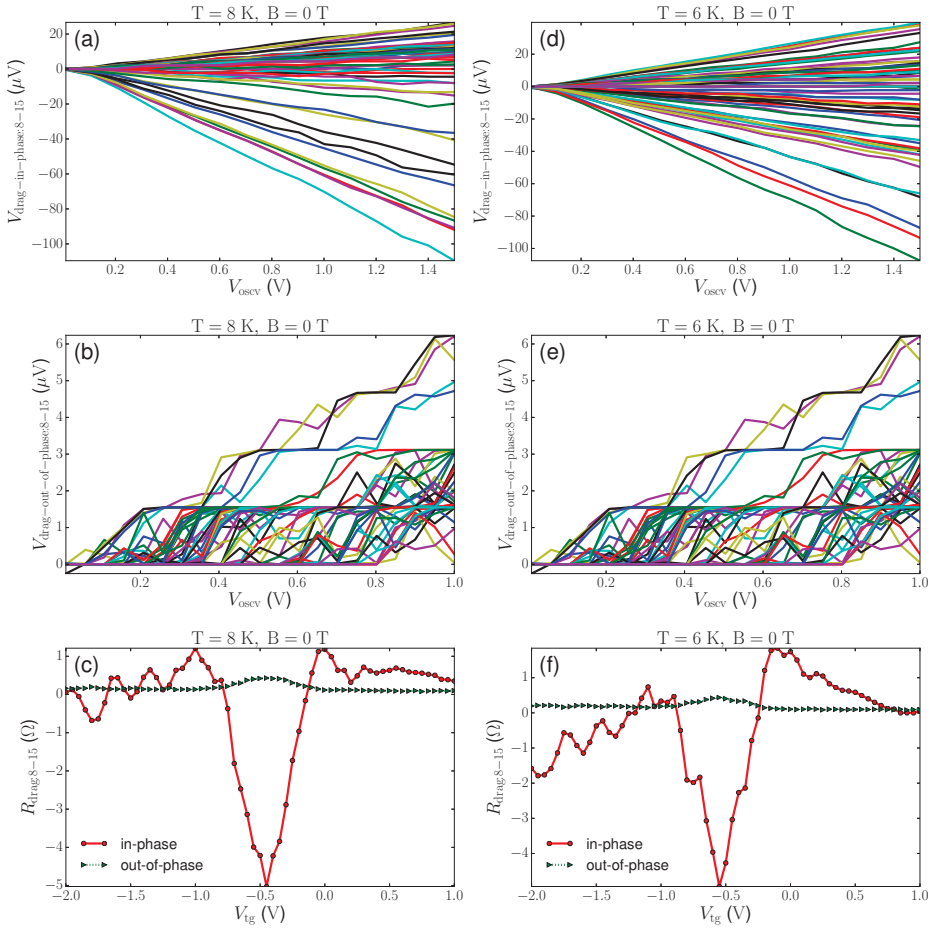

Fig. F.4. In-phase (a,d) and out-of-phase (b,e) components of the Coulomb drag voltage between graphene and $NbSe_2$ vs. V_{oscv} and the drag resistance (c,f) vs. V_{tg} at two different temperatures of 8 K (a-c) and 6 K (d-f).

G. DERIVATIONS FOR BOLTZMANN EQUATIONS

G.1 Linearizion of Boltzmann equation

We assume every layer and the applied electrical field within it are uniform. We can then drop the position dependence in f_μ, S_μ and $S_{\mu\nu}$. Furthermore, we assume the electrical fields are the only force to drive the distribution out of equilibrium and they are small and so is the derivation of f_μ from the equilibrium distribution function n_μ which is

$$n_\mu(\mathbf{k}_\mu) = \left[\exp\left(\frac{e_\mu(\mathbf{k}_\mu) - \zeta_\mu}{k_\mathrm{B}T_\mu}\right) + 1\right]^{-1} \tag{G.1}$$

where k_B is the Boltzmann constant, T_μ is the temperature and ζ_μ is the chemical potential. Therefore the linear response theory can be applied to calculate the transport coefficients. We can write

$$f_\mu = n_\mu(1 - \bar{n}_\mu\phi_\mu) \tag{G.2}$$

so that $\phi_\mu \propto \sum_\nu c_{\mu\nu}\mathbf{F}_\nu$ where $c_{\mu\nu}$ are constants (may depends on $\mathbf{k}_\mu$s, but this is not an essential point). Up to the first order in ϕ_μ, the collision integrals are linearized as

$$\begin{aligned} C_\mu[f_\mu](\mathbf{k}_\mu) = \int_{B_\mu} [Q_\mathrm{c}(\mu'\mu) - Q_\mathrm{c}(\mu\mu') - \phi'_\mu\{Q_\mathrm{c}(\mu'\mu)\bar{n}'_\mu + Q_\mathrm{c}(\mu\mu')n'_\mu\} \\ + \phi_\mu\{Q_\mathrm{c}(\mu'\mu)n_\mu + Q_\mathrm{c}(\mu\mu')\bar{n}_\mu\}]\mathrm{d}\mu' \end{aligned} \tag{G.3}$$

and

$$\begin{aligned} C_{\mu\nu}[f_\mu, f_\nu](\mathbf{k}_\mu) = \int_{B_\mu B_\nu^2} & [Q_{\rm cc}(\mu'\mu\nu'\nu) - Q_{\rm cc}(\mu\mu'\nu\nu') \\ & - \phi'_\mu\{Q_{\rm cc}(\mu'\mu\nu'\nu)\bar{n}'_\mu + Q_{\rm cc}(\mu\mu'\nu\nu')n'_\mu\} \\ & + \phi_\mu\{Q_{\rm cc}(\mu'\mu\nu'\nu)n_\mu + Q_{\rm cc}(\mu\mu'\nu\nu')\bar{n}_\mu\} \\ & - \phi'_\nu\{Q_{\rm cc}(\mu'\mu\nu'\nu)\bar{n}'_\nu + Q_{\rm cc}(\mu\mu'\nu\nu')n'_\nu\} \\ & + \phi_\nu\{Q_{\rm cc}(\mu'\mu\nu'\nu)n_\nu + Q_{\rm cc}(\mu\mu'\nu\nu')\bar{n}_\nu\}]{\rm d}\mu'{\rm d}\nu{\rm d}\nu', \end{aligned} \tag{G.4}$$

where we have used the following shorthand notations

$$\begin{aligned} Q_{\rm c}(\mu'\mu) &\equiv S_{\rm c}(\mathbf{k}'_\mu, \mathbf{k}_\mu)n'_\mu\bar{n}_\mu, \\ Q_{\rm cc}(\mu'\mu\nu'\nu) &\equiv S_{\rm cc}(\mathbf{k}'_\mu, \mathbf{k}_\mu, \mathbf{k}'_\nu, \mathbf{k}_\nu)n'_\mu\bar{n}_\mu n'_\nu\bar{n}_\nu, \\ {\rm d}\mu' &\equiv {\rm d}\mathbf{k}'_\mu,\ {\rm d}\nu \equiv {\rm d}\mathbf{k}_\nu. \end{aligned} \tag{G.5}$$

The principle of detailed balance requires

$$\begin{aligned} Q_{\rm c}(\mu'\mu) &= Q_{\rm c}(\mu\mu'), \\ Q_{\rm cc}(\mu'\mu\nu'\nu) &= Q_{\rm cc}(\mu\mu'\nu\nu'). \end{aligned} \tag{G.6}$$

The binary collision probability $S_{\rm cc}(\mathbf{k}'_\mu, \mathbf{k}_\mu, \mathbf{k}'_\nu, \mathbf{k}_\nu)$ should be symmetrical with respect to exchanging the layer indices μ and ν, therefore we have another symmetry property for $Q_{\rm cc}$:

$$Q_{\rm cc}(\mu'\mu\nu'\nu) = Q_{\rm cc}(\nu'\nu\mu'\mu). \tag{G.7}$$

Due to the time reversal invariance of the Coulomb scattering, the transition probability has the following symmetry property

$$S_{\rm cc}(\mathbf{k}'_\mu, \mathbf{k}_\mu, \mathbf{k}'_\nu, \mathbf{k}_\nu) = S_{\rm cc}(\mathbf{k}_\mu, \mathbf{k}'_\mu, \mathbf{k}_\nu, \mathbf{k}'_\nu). \tag{G.8}$$

Note that there is no similar relation for the single carrier scattering transition probability. Then the second equation in Eq. (G.6) implies

$$\frac{e_\mu(\mathbf{k}_\mu)}{k_{\rm B}T_\mu} + \frac{e_\nu(\mathbf{k}_\nu)}{k_{\rm B}T_\nu} = \frac{e_\mu(\mathbf{k}'_\mu)}{k_{\rm B}T_\mu} + \frac{e_\nu(\mathbf{k}'_\nu)}{k_{\rm B}T_\nu}. \tag{G.9}$$

The kinetic energy is conserved during the collision, i.e.,

$$e_\mu(\mathbf{k}_\mu) + e_\nu(\mathbf{k}_\nu) = e_\mu(\mathbf{k}'_\mu) + e_\nu(\mathbf{k}'_\nu). \tag{G.10}$$

Eq. (G.9) and Eq. (G.10) immediately yield $T_\mu = T_\nu$. This can also be obtained from the following arguments. Since

$$\frac{Q_{\rm cc}(\mu'\mu\nu'\nu)}{Q_{\rm cc}(\mu\mu'\nu\nu')} = \exp\left((e_\mu(\mathbf{k}_\mu) - e_\mu(\mathbf{k}'_\mu))\left(\frac{1}{T_\mu} - \frac{1}{T_\nu}\right)\right), \tag{G.11}$$

if $T_\mu > T_\nu$, the above ratio is larger than 1 if $e_\mu(\mathbf{k}_\mu) < e_\mu(\mathbf{k}'_\mu)$, which means the energy transfer due to interlayer Coulomb scatterings from layer μ to layer ν is easier until the temperatures are the same. In the following, we thus assign $T_\mu = T$ for every layer.

The collision integrals can now be written in the following concise form

$$\begin{aligned} C_\mu[\phi_\mu](\mathbf{k}_\mu) &= \int_{B_\mu} Q_{\rm c}(\mu'\mu)[\phi_\mu - \phi'_\mu]{\rm d}\mu', \\ C_{\mu\nu}[\phi_\mu, \phi_\nu](\mathbf{k}_\mu) &= \int_{B_\mu B_\nu^2} Q_{\rm cc}(\mu'\mu\nu'\nu)[\phi_\mu - \phi'_\mu + \phi_\nu - \phi'_\nu]{\rm d}\mu'{\rm d}\nu{\rm d}\nu'. \end{aligned} \tag{G.12}$$

Note that in these integrals and the following, ν, ν', ϕ_ν and ϕ'_ν have to be replaced by, e.g., μ'', μ''', ϕ''_μ and ϕ'''_μ respectively, when $\nu = \mu$. It is of course understood that $\phi''_\mu = \phi_\mu(\mathbf{k}''_\mu)$ and similarly for ϕ'''_μ.

G.2 Entropy production

We start from the original Boltzmann equation, Eq. (5.38), to derive the continuity equation for the entropy density s_μ and entropy flux $\mathbf{J}_\mu$ that are defined as

$$\begin{aligned} s_\mu(\mathbf{x}) &\equiv -k_{\rm B}\int_{B_\mu} (f_\mu \ln f_\mu + \bar{f}_\mu \ln \bar{f}_\mu){\rm d}\mu, \\ \mathbf{J}_\mu(\mathbf{x}) &\equiv -k_{\rm B}\int_{B_\mu} \mathbf{v}_\mu(\mathbf{k}_\mu)(f_\mu \ln f_\mu + \bar{f}_\mu \ln \bar{f}_\mu){\rm d}\mu. \end{aligned} \tag{G.13}$$

It can be shown that[1]

$$\begin{aligned}
\partial_t s_\mu + \nabla_{\mathbf{x}} \cdot \mathbf{J}_\mu &= -k_{\mathrm{B}} \int_{B_\mu} \left(\partial_t f_\mu + \mathbf{v}_\mu \cdot \nabla_{\mathbf{x}} f_\mu + \frac{\mathbf{F}_\mu}{\hbar} \cdot \nabla_{\mathbf{k}} f_\mu \right) (\ln f_\mu - \ln \bar{f}_\mu) \mathrm{d}\mu \\
&= -k_{\mathrm{B}} \int_{B_\mu} \left(C_\mu[f_\mu] + \sum_\nu C_{\mu\nu}[f_\mu, f_\nu] \right) (\ln f_\mu - \ln \bar{f}_\mu) \mathrm{d}\mu \\
&= -k_{\mathrm{B}} \int_{B_\mu} \left(C_\mu[f_\mu] + \sum_\nu C_{\mu\nu}[f_\mu, f_\nu] \right) \left(-\frac{e_\mu - \zeta_\mu}{k_{\mathrm{B}} T} - \phi_\mu \right) \mathrm{d}\mu \\
&= k_{\mathrm{B}} \int_{B_\mu} \left(C_\mu[f_\mu] + \sum_\nu C_{\mu\nu}[f_\mu, f_\nu] \right) \phi_\mu \mathrm{d}\mu \\
&= k_{\mathrm{B}} \langle \phi_\mu, X_\mu \rangle = \frac{\mathbf{j}_\mu \cdot \mathbf{E}_\mu}{T}
\end{aligned} \tag{G.14}$$

where we have used the assumptions that all the scatterings conserves the kinetic energy and carrier number. At steady state for homogeneous systems, the LHS of Eq. (G.14) is apparently zero but its last line can be nonzero. This paradox stems from our reckless usage of the quantity such as the entropy density at equilibrium in the nonequilibrium transport processes. Nonetheless, it is a insightful result.

[1]We have used $\nabla_{\mathbf{x}} \cdot \mathbf{v}_\mu \sim \nabla_{\mathbf{x}} \times \mathbf{F}_\mu = 0$ and

$$\int_{B_\mu} (\nabla_{\mathbf{k}} f_\mu)(\ln f_\mu - \ln \bar{f}_\mu) \mathrm{d}\mu = \int_{B_\mu} \nabla_{\mathbf{k}} (f_\mu \ln f_\mu + \bar{f}_\mu \ln \bar{f}_\mu) \mathrm{d}\mu = 0,$$

since f_μ is periodic in B_μ.

G.3 Theory of interlayer excitons

We begin with the general form of the single-particle Hamiltonian[2]

$$H_\mu = -\frac{\hbar^2 \nabla^2_{\mathbf{r}_\mu}}{2m_\mu} + V_\mu(\mathbf{r}_\mu) \tag{G.15}$$

for carriers in each individual layer, where m_μ is the bare mass of the carriers. The periodicity of $V_\mu(\mathbf{r}_\mu)$, i.e., $V_\mu(\mathbf{r}_\mu) = V_\mu(\mathbf{r}_\mu + \mathbf{R}_\mu)$ where $\mathbf{R}_\mu = \sum_i s^i_\mu \mathbf{a}^i_\mu$ for the base vectors $\mathbf{a}^i_\mu$ and any integers s^i_μ (i is the spatial indices), leads to the band structure of e_μ. For carriers in the layer μ and ν, if the interlayer interaction $V_{\mu\nu}$ is included, the Hamiltonian can be written as

$$H_{\mu\nu} = H_\mu + H_\nu + V(\mathbf{r}_\mu - \mathbf{r}_\nu) = -\frac{\hbar^2 \nabla^2_{\mathbf{r}_\mu}}{2m_\mu} - \frac{\hbar^2 \nabla^2_{\mathbf{r}_\nu}}{2m_\nu} + V_\mu(\mathbf{r}_\mu) + V_\nu(\mathbf{r}_\nu) + V_{\mu\nu}(\bar{\mathbf{r}}_{\mu\nu}), \tag{G.16}$$

where $\bar{\mathbf{r}}_{\mu\nu} \equiv \mathbf{r}_\mu - \mathbf{r}_\nu$ is the relative coordinate. If transform to the center of mass coordinate by writing

$$\mathbf{r}_{\mu\nu} = \frac{m_\mu \mathbf{r}_\mu + m_\nu \mathbf{r}_\nu}{m_\mu + m_\nu}, \tag{G.17}$$

the Hamiltonian now becomes

$$H_{\mu\nu} = -\frac{\hbar^2 \nabla^2_{\mathbf{r}_{\mu\nu}}}{2(m_\mu + m_\nu)} - \frac{\hbar^2 \nabla^2_{\bar{\mathbf{r}}_{\mu\nu}}}{2m_{\mu\nu}} + V_\mu\left(\mathbf{r}_{\mu\nu} + \frac{m_{\mu\nu}}{m_\mu}\bar{\mathbf{r}}_{\mu\nu}\right) + V_\nu\left(\mathbf{r}_{\mu\nu} - \frac{m_{\mu\nu}}{m_\nu}\bar{\mathbf{r}}_{\mu\nu}\right) + V_{\mu\nu}(\bar{\mathbf{r}}_{\mu\nu}), \tag{G.18}$$

where $m_{\mu\nu} \equiv \frac{m_\mu m_\nu}{m_\mu + m_\nu}$. Apparently $V_{\mu\nu}(\bar{\mathbf{r}}_{\mu\nu})$ has no long range order. If the potential sum of $V_\mu + V_\nu$ has no long range order for $\mathbf{r}_{\mu\nu}$, i.e., the Fourier transform of its spatial correlation function has no regular peaks, then the energy band theory for the center

[2]For holes defined in the semiconductor community, similar Hamiltonian can also be written down. The bare mass can be taken as the same as that of electrons. The holes possesses positive charge, which is not encoded in the Hamiltonian (neither for electrons), since charge as a quantum number has to be dealt with in the theory of quantum fields explicitly.

of mass motion is not applicable. We thus assume there exists long range order for $\mathbf{r}_{\mu\nu}$, i.e., for some nonzero $\mathbf{R}_{\mu\nu}$[3] we have

$$V(\mathbf{r}_{\mu\nu}, \bar{\mathbf{r}}_{\mu\nu}) = V(\mathbf{r}_{\mu\nu} + \mathbf{R}_{\mu\nu}, \bar{\mathbf{r}}_{\mu\nu}), \tag{G.19}$$

for any $\bar{\mathbf{r}}_{\mu\nu}$, where

$$V(\mathbf{r}_{\mu\nu}, \bar{\mathbf{r}}_{\mu\nu}) \equiv V_\mu \left(\mathbf{r}_{\mu\nu} + \frac{m_{\mu\nu}}{m_\mu} \bar{\mathbf{r}}_{\mu\nu} \right) + V_\nu \left(\mathbf{r}_{\mu\nu} - \frac{m_{\mu\nu}}{m_\nu} \bar{\mathbf{r}}_{\mu\nu} \right). \tag{G.20}$$

This requires that the translation of $\mathbf{r}_{\mu\nu}$ also makes V_μ and V_ν invariant, which necessarily desires

$$\sum_i s_\mu^i \mathbf{a}_\mu^i = \sum_i s_\nu^i \mathbf{a}_\nu^i \neq 0, \tag{G.21}$$

for some integers s_μ^i and s_ν^i. The lattices in layer μ and ν are then commensurate. Now the center of mass motion has energy band $E_{n\mathbf{k}_{\mu\nu}}(\bar{\mathbf{r}}_{\mu\nu})$ with Bloch state $\psi_{n\mathbf{k}_{\mu\nu}}(\mathbf{r}_{\mu\nu}, \bar{\mathbf{r}}_{\mu\nu})$, band index n and quasi-momentum $\mathbf{k}_{\mu\nu}$. Note that the relative motion coordinate $\bar{\mathbf{r}}_{\mu\nu}$ here is considered as a parameter.

Now, the most general form of the two body wave function can be written as

$$\Phi_{\mu\nu} = \sum_{n\mathbf{k}_{\mu\nu}} \psi_{n\mathbf{k}_{\mu\nu}}(\mathbf{r}_{\mu\nu}, \bar{\mathbf{r}}_{\mu\nu}) \phi_{n\mathbf{k}_{\mu\nu}}(\bar{\mathbf{r}}_{\mu\nu}), \tag{G.22}$$

since the Bloch states $\psi_{n\mathbf{k}_{\mu\nu}}(\mathbf{r}_{\mu\nu}, \bar{\mathbf{r}}_{\mu\nu})$ form a complete set for any $\bar{\mathbf{r}}_{\mu\nu}$. By plugging this into the equation $H_{\mu\nu}\Phi_{\mu\nu} = e_{\mu\nu}\Phi_{\mu\nu}$, we have the following matrix form of equation

$$\left[\frac{\left[-i\hbar \nabla_{\bar{\mathbf{r}}_{\mu\nu}} - \mathbf{A}(\bar{\mathbf{r}}_{\mu\nu}) \right]^2}{2m_{\mu\nu}} + V_{\mu\nu}(\bar{\mathbf{r}}_{\mu\nu}) + W(\bar{\mathbf{r}}_{\mu\nu}) \right] \phi(\bar{\mathbf{r}}_{\mu\nu}) = e_{\mu\nu} \phi(\bar{\mathbf{r}}_{\mu\nu}), \tag{G.23}$$

where $\phi(\bar{\mathbf{r}}_{\mu\nu})$ is the vector with components of $\phi_{n\mathbf{k}_{\mu\nu}}(\bar{\mathbf{r}}_{\mu\nu})$, and

$$\begin{aligned} \mathbf{A}(\bar{\mathbf{r}}_{\mu\nu}) &\equiv i\hbar \int \psi^*(\mathbf{r}_{\mu\nu}, \bar{\mathbf{r}}_{\mu\nu}) \otimes \nabla_{\bar{\mathbf{r}}_{\mu\nu}} \psi(\mathbf{r}_{\mu\nu}, \bar{\mathbf{r}}_{\mu\nu}) \mathrm{d}\mathbf{r}_{\mu\nu}, \\ W(\bar{\mathbf{r}}_{\mu\nu}) &\equiv \frac{\hbar^2}{2m_{\mu\nu}} \int \nabla_{\bar{\mathbf{r}}_{\mu\nu}} \psi^* \otimes \nabla_{\bar{\mathbf{r}}_{\mu\nu}} \psi \mathrm{d}\mathbf{r}_{\mu\nu} - \frac{\mathbf{A}^2(\bar{\mathbf{r}}_{\mu\nu})}{2m_{\mu\nu}} + \mathrm{diag}\{E_{n\mathbf{k}_{\mu\nu}}(\bar{\mathbf{r}}_{\mu\nu})\}. \end{aligned} \tag{G.24}$$

[3]The dimension of the space spanned by these vectors could be less or same as that by $\mathbf{a}_\mu^i$ or $\mathbf{a}_\nu^i$.

Here diag$\{E_{n\mathbf{k}_{\mu\nu}}(\bar{\mathbf{r}}_{\mu\nu})\}$ is a diagonal matrix with diagonal elements of $E_{n\mathbf{k}_{\mu\nu}}(\bar{\mathbf{r}}_{\mu\nu})$, and ψ is a vector with components $\psi_{n\mathbf{k}_{\mu\nu}}$. After the following unitary transformation

$$\psi(\mathbf{r}_{\mu\nu}, \bar{\mathbf{r}}_{\mu\nu}) \rightarrow U\psi(\mathbf{r}_{\mu\nu}, \bar{\mathbf{r}}_{\mu\nu}), \tag{G.25}$$

we have

$$\mathbf{A}(\bar{\mathbf{r}}_{\mu\nu}) \rightarrow U^*\mathbf{A}(\bar{\mathbf{r}}_{\mu\nu})U^T + i\hbar U^*\nabla_{\bar{\mathbf{r}}_{\mu\nu}}U^T, \;\; W(\bar{\mathbf{r}}_{\mu\nu}) \rightarrow U^*W(\bar{\mathbf{r}}_{\mu\nu})U^T. \tag{G.26}$$

Here $\mathbf{A}$ is actually the Berry vector potential. If the curl of $\mathbf{A}$ is zero, it can be written as the gradient of a matrix $B(\bar{\mathbf{r}}_{\mu\nu})$, i.e., $\mathbf{A}(\bar{\mathbf{r}}_{\mu\nu}) = \nabla_{\bar{\mathbf{r}}_{\mu\nu}}B(\bar{\mathbf{r}}_{\mu\nu})$. In addition, if B can be written as a sum of commuting Hermitian matrices, the same set of matrices can be used to expand a matrix $\theta(\bar{\mathbf{r}}_{\mu\nu})$ such that $U = e^{i\theta}$. By letting $\theta = B^T/\hbar$, the matrix $\mathbf{A}$ is then transformed to zero. In general, either the curl of $\mathbf{A}$ is nonzero, or $\mathbf{A}$ can not be expanded as a sum of commuting Hermitian matrices, thus $\mathbf{A}$ cannot be transformed to zero.

Now the square bracket in Eq. (G.23) is Hermitian, and it permits real values of $e_{\mu\nu}$. In general, $\mathbf{A}$ and W cannot be diagonalized simultaneously using a matrix independent of $\bar{\mathbf{r}}_{\mu\nu}$, thus Eq. (G.23) cannot be decoupled into a set of scalar equations.

We are seeking for excitons of Bloch-like, i.e.,

$$\Phi_{\mu\nu}(\mathbf{r}_{\mu\nu} + \mathbf{R}_{\mu\nu}, \bar{\mathbf{r}}_{\mu\nu}) = e^{i\mathbf{K}\cdot\mathbf{R}_{\mu\nu}}\Phi_{\mu\nu}(\mathbf{r}_{\mu\nu}, \bar{\mathbf{r}}_{\mu\nu}), \tag{G.27}$$

for some $\mathbf{K}$. We then arrives at

$$e^{i\mathbf{K}\cdot\mathbf{R}_{\mu\nu}}\phi_{n\mathbf{k}_{\mu\nu}}(\bar{\mathbf{r}}_{\mu\nu}) = e^{i\mathbf{k}_{\mu\nu}\cdot\mathbf{R}_{\mu\nu}}\phi_{n\mathbf{k}_{\mu\nu}}(\bar{\mathbf{r}}_{\mu\nu}). \tag{G.28}$$

Thus every $\mathbf{k}_{\mu\nu}$ can only be different from $\mathbf{K}$ by some reciprocal vector $\mathbf{G}_{\mu\nu}$. We can let $\mathbf{K} = \mathbf{k}_{\mu\nu}$. The Bloch-like exciton wave function then is

$$\begin{aligned}\Phi_{\mathbf{k}_{\mu\nu}}(\mathbf{r}_{\mu\nu}, \bar{\mathbf{r}}_{\mu\nu}) &= \sum_{n\mathbf{G}_{\mu\nu}} \psi_{n\mathbf{k}_{\mu\nu}+\mathbf{G}_{\mu\nu}}(\mathbf{r}_{\mu\nu}, \bar{\mathbf{r}}_{\mu\nu})\phi_{n\mathbf{k}_{\mu\nu}+\mathbf{G}_{\mu\nu}}(\bar{\mathbf{r}}_{\mu\nu}) \\ &= \sum_{n} \psi_{n\mathbf{k}_{\mu\nu}}(\mathbf{r}_{\mu\nu}, \bar{\mathbf{r}}_{\mu\nu}) \sum_{\mathbf{G}_{\mu\nu}} \phi_{n\mathbf{k}_{\mu\nu}+\mathbf{G}_{\mu\nu}}(\bar{\mathbf{r}}_{\mu\nu}) \\ &= \sum_{n} \psi_{n\mathbf{k}_{\mu\nu}}(\mathbf{r}_{\mu\nu}, \bar{\mathbf{r}}_{\mu\nu})\varphi_{n\mathbf{k}_{\mu\nu}}(\bar{\mathbf{r}}_{\mu\nu}),\end{aligned} \tag{G.29}$$

where $\varphi_{n\mathbf{k}_{\mu\nu}}(\bar{\mathbf{r}}_{\mu\nu}) \equiv \sum_{\mathbf{G}_{\mu\nu}} \phi_{n\mathbf{k}_{\mu\nu}+\mathbf{G}_{\mu\nu}}(\bar{\mathbf{r}}_{\mu\nu})$. Therefore, the Bloch-like excitons can only be formed from the center of mass wave functions from different bands with the same wave vector. If we require $\mathbf{k}_{\mu\nu}$ to be in the reduced Brillouin zone, we simply have $\varphi_{n\mathbf{k}_{\mu\nu}}(\bar{\mathbf{r}}_{\mu\nu}) = \phi_{n\mathbf{k}_{\mu\nu}}(\bar{\mathbf{r}}_{\mu\nu})$. In order to require that the eigenstates of $H_{\mu\nu}$ are Bloch-like, it suffices to require $\mathbf{A}$ and W to be diagonal with respect to $\mathbf{k}_{\mu\nu}$, i.e.,

$$\begin{aligned}\int \psi^*_{n'\mathbf{k}'_{\mu\nu}}(\mathbf{r}_{\mu\nu}, \bar{\mathbf{r}}_{\mu\nu})\nabla_{\bar{\mathbf{r}}_{\mu\nu}}\psi_{n\mathbf{k}_{\mu\nu}}(\mathbf{r}_{\mu\nu}, \bar{\mathbf{r}}_{\mu\nu})\mathrm{d}\mathbf{r}_{\mu\nu} &\propto \delta_{\mathbf{k}'_{\mu\nu}\mathbf{k}_{\mu\nu}}, \\ \int \nabla_{\bar{\mathbf{r}}_{\mu\nu}}\psi^*_{n'\mathbf{k}'_{\mu\nu}}(\mathbf{r}_{\mu\nu}, \bar{\mathbf{r}}_{\mu\nu}) \cdot \nabla_{\bar{\mathbf{r}}_{\mu\nu}}\psi_{n\mathbf{k}_{\mu\nu}}(\mathbf{r}_{\mu\nu}, \bar{\mathbf{r}}_{\mu\nu})\mathrm{d}\mathbf{r}_{\mu\nu} &\propto \delta_{\mathbf{k}'_{\mu\nu}\mathbf{k}_{\mu\nu}},\end{aligned} \tag{G.30}$$

which can be proved as the following. The Bloch state $\psi_{n\mathbf{k}_{\mu\nu}}(\mathbf{r}_{\mu\nu}, \bar{\mathbf{r}}_{\mu\nu})$ can be written as

$$\psi_{n\mathbf{k}_{\mu\nu}}(\mathbf{r}_{\mu\nu}, \bar{\mathbf{r}}_{\mu\nu}) = e^{i\mathbf{k}_{\mu\nu}\cdot\mathbf{r}_{\mu\nu}} \sum_{\mathbf{G}_{\mu\nu}} c_{n\mathbf{k}_{\mu\nu}}(\mathbf{G}_{\mu\nu}, \bar{\mathbf{r}}_{\mu\nu})e^{i\mathbf{G}_{\mu\nu}\cdot\mathbf{r}_{\mu\nu}}. \tag{G.31}$$

Then

$$\begin{aligned}\int \psi^*_{n'\mathbf{k}'_{\mu\nu}}\nabla_{\bar{\mathbf{r}}_{\mu\nu}}\psi_{n\mathbf{k}_{\mu\nu}}\mathrm{d}\mathbf{r}_{\mu\nu} = &\sum_{\mathbf{G}_{\mu\nu}\mathbf{G}'_{\mu\nu}} c^*_{n'\mathbf{k}'_{\mu\nu}}(\mathbf{G}'_{\mu\nu}, \bar{\mathbf{r}}_{\mu\nu})\nabla_{\bar{\mathbf{r}}_{\mu\nu}}c_{n\mathbf{k}_{\mu\nu}}(\mathbf{G}_{\mu\nu}, \bar{\mathbf{r}}_{\mu\nu}) \\ &\times \int e^{i(\mathbf{k}_{\mu\nu}-\mathbf{k}'_{\mu\nu}+\mathbf{G}_{\mu\nu}-\mathbf{G}'_{\mu\nu})\cdot\mathbf{r}_{\mu\nu}} \propto \delta_{\mathbf{k}'_{\mu\nu}\mathbf{k}_{\mu\nu}}.\end{aligned} \tag{G.32}$$

A similar proof can be provided for the second line in Eq. (G.30). This trick however cannot be used for the band indices. Nevertheless, we can assume that only one

relevant band is considered in each layer and the other bands are far away from it so that $\mathbf{A}$ and W are diagonal with respect to the band index n. We can now drop the band index.

For the ansatz $\Phi_{\mathbf{k}_{\mu\nu}} = \psi_{\mathbf{k}_{\mu\nu}}(\mathbf{r}_{\mu\nu}, \bar{\mathbf{r}}_{\mu\nu})\phi_{\mathbf{k}_{\mu\nu}}(\bar{\mathbf{r}}_{\mu\nu})$ with energy $e_{\mu\nu}(\mathbf{k}_{\mu\nu})$, the relative motion reduced from Eq. (G.23) is

$$\left[\frac{\left[-i\hbar\nabla_{\bar{\mathbf{r}}_{\mu\nu}} - \mathbf{A}_{\mathbf{k}_{\mu\nu}}(\bar{\mathbf{r}}_{\mu\nu})\right]^2}{2m_{\mu\nu}} + V_{\mu\nu}(\bar{\mathbf{r}}_{\mu\nu}) + W_{\mathbf{k}_{\mu\nu}}(\bar{\mathbf{r}}_{\mu\nu})\right]\phi_{\mathbf{k}_{\mu\nu}}(\bar{\mathbf{r}}_{\mu\nu}) = e_{\mu\nu}\phi_{\mathbf{k}_{\mu\nu}}(\bar{\mathbf{r}}_{\mu\nu}), \tag{G.33}$$

where

$$\begin{aligned}
\mathbf{A}_{\mathbf{k}_{\mu\nu}}(\bar{\mathbf{r}}_{\mu\nu}) &\equiv i\hbar \int \psi^*_{\mathbf{k}_{\mu\nu}}(\mathbf{r}_{\mu\nu}, \bar{\mathbf{r}}_{\mu\nu})\nabla_{\bar{\mathbf{r}}_{\mu\nu}}\psi_{\mathbf{k}_{\mu\nu}}(\mathbf{r}_{\mu\nu}, \bar{\mathbf{r}}_{\mu\nu})\mathrm{d}\mathbf{r}_{\mu\nu}, \\
W_{\mathbf{k}_{\mu\nu}}(\bar{\mathbf{r}}_{\mu\nu}) &\equiv \frac{\hbar^2}{2m_{\mu\nu}} \int \nabla_{\bar{\mathbf{r}}_{\mu\nu}}\psi^*_{\mathbf{k}_{\mu\nu}} \cdot \nabla_{\bar{\mathbf{r}}_{\mu\nu}}\psi_{\mathbf{k}_{\mu\nu}}\mathrm{d}\mathbf{r}_{\mu\nu} - \frac{\mathbf{A}^2_{\mathbf{k}_{\mu\nu}}(\bar{\mathbf{r}}_{\mu\nu})}{2m_{\mu\nu}} + E_{\mathbf{k}_{\mu\nu}}(\bar{\mathbf{r}}_{\mu\nu}),
\end{aligned} \tag{G.34}$$

For the gauge transformation

$$\psi_{\mathbf{k}_{\mu\nu}}(\mathbf{r}_{\mu\nu}, \bar{\mathbf{r}}_{\mu\nu}) \to e^{i\theta(\bar{\mathbf{r}}_{\mu\nu})}\psi_{\mathbf{k}_{\mu\nu}}(\mathbf{r}_{\mu\nu}, \bar{\mathbf{r}}_{\mu\nu}), \tag{G.35}$$

we have $\mathbf{A}_{\mathbf{k}_{\mu\nu}}(\bar{\mathbf{r}}_{\mu\nu}) \to \mathbf{A}_{\mathbf{k}_{\mu\nu}}(\bar{\mathbf{r}}_{\mu\nu}) - \hbar\nabla_{\bar{\mathbf{r}}_{\mu\nu}}\theta(\bar{\mathbf{r}}_{\mu\nu})$[4], while $W_{\mathbf{k}_{\mu\nu}}$ remains gauge invariant. Therefore, the relative motion is now described by the effective vector potential $\mathbf{A}_{\mathbf{k}_{\mu\nu}}$ and scalar potential $V_{\mu\nu} + W_{\mathbf{k}_{\mu\nu}}$.

The symmetry properties of $\Phi_{\mathbf{k}_{\mu\nu}}$ and $e_{\mu\nu}(\mathbf{k}_{\mu\nu})$ (e.g., translations in the real and vector spaces) are apparent. The excitonic states can also be expanded using the Bloch states $\psi^{\mu}_{\mathbf{k}_{\mu}}(\mathbf{r}_{\mu})$ with energy $e^{\mu}(\mathbf{k}_{\mu})$ from individual layers,

$$\Phi_{\mathbf{k}_{\mu\nu}}(\mathbf{r}_{\mu\nu}, \bar{\mathbf{r}}_{\mu\nu}) = \int_{B_\mu B_\nu} F(\mathbf{k}_{\mu}, \mathbf{k}_{\nu})\psi^{\mu}_{\mathbf{k}_{\mu}}(\mathbf{r}_{\mu})\psi^{\nu}_{\mathbf{k}_{\nu}}(\mathbf{r}_{\nu})\mathrm{d}\mathbf{k}_{\mu}\mathrm{d}\mathbf{k}_{\nu}, \tag{G.36}$$

[4]In general, $\mathbf{A}_{\mathbf{k}_{\mu\nu}}$ cannot be gauged away if both the real and complex components of $\psi_{\mathbf{k}_{\mu\nu}}$ are nonzero. It can only be gauged away if its curl $\nabla_{\bar{\mathbf{r}}_{\mu\nu}} \times \mathbf{A}_{\mathbf{k}_{\mu\nu}}(\bar{\mathbf{r}}_{\mu\nu})$ is zero.

where B_μ is the reduced (e.g., first) Brillouin zone for layer μ. The translation symmetry requires

$$\begin{aligned}\Phi_{\mathbf{k}_{\mu\nu}}(\mathbf{r}_{\mu\nu}+\mathbf{R}_{\mu\nu},\bar{\mathbf{r}}_{\mu\nu}) &= \int_{B_\mu B_\nu} F(\mathbf{k}_\mu,\mathbf{k}_\nu)\psi^\mu_{\mathbf{k}_\mu}(\mathbf{r}_\mu)\psi^\nu_{\mathbf{k}_\nu}(\mathbf{r}_\nu)e^{i(\mathbf{k}_\mu+\mathbf{k}_\nu)\cdot\mathbf{R}_{\mu\nu}}\mathrm{d}\mathbf{k}_\mu\mathrm{d}\mathbf{k}_\nu \\ &= e^{i\mathbf{k}_{\mu\nu}\cdot\mathbf{R}_{\mu\nu}}\Phi_{\mathbf{k}_{\mu\nu}}(\mathbf{r}_{\mu\nu},\bar{\mathbf{r}}_{\mu\nu}).\end{aligned} \tag{G.37}$$

Therefore, for nonzero value of $F(\mathbf{k}_\mu,\mathbf{k}_\nu)$ we must have

$$\mathbf{k}_\mu+\mathbf{k}_\nu=\mathbf{k}_{\mu\nu}, \tag{G.38}$$

where $\mathbf{k}_{\mu\nu}$ is allowed to be out side of the reduced Brillouin zone $B_{\mu\nu}$. This indicates that only the carriers with total quasi-momentum $\mathbf{k}_{\mu\nu}$ can be allowed to compose the Bloch-like exciton states, an manifestation of the conservation of the quasi-momentum during the formation of the composite particle of the exciton. It is therefore convenient to write the exciton Bloch state as

$$\Phi_{\mathbf{k}_{\mu\nu}}(\mathbf{r}_{\mu\nu},\bar{\mathbf{r}}_{\mu\nu})=\int F_{\mathbf{k}_{\mu\nu}}(\mathbf{k})\psi^\mu_{\mathbf{k}_{\mu\nu}/2+\mathbf{k}}(\mathbf{r}_\mu)\psi^\nu_{\mathbf{k}_{\mu\nu}/2-\mathbf{k}}(\mathbf{r}_\nu)\mathrm{d}\mathbf{k}. \tag{G.39}$$

The integration range is over the $\mathbf{k}$ space such that $\mathbf{k}_{\mu\nu}/2+\mathbf{k}\in B_\mu$ and $\mathbf{k}_{\mu\nu}/2-\mathbf{k}\in B_\nu$. If such $\mathbf{k}$ does not exist, $\mathbf{k}_{\mu\nu}$ must be outside of $B_{\mu\nu}$ and it can always be inside $B_{\mu\nu}$ by an appropriate translation through a reciprocal lattice vector, since the volume of $B_{\mu\nu}$ is smaller or equal to the minimum of the volumes of B_μ and B_ν. The orthonormality of $\Phi_{\mathbf{k}_{\mu\nu}}(\mathbf{r}_{\mu\nu},\bar{\mathbf{r}}_{\mu\nu})$ requires

$$\int F^*_{\mathbf{k}_{\mu\nu}}(\mathbf{k})F_{\mathbf{k}_{\mu\nu}}(\mathbf{k})\mathrm{d}\mathbf{k}=1. \tag{G.40}$$

We can introduce the creation (annihilation) operator $a^{\mu\dagger}_{\mathbf{k}_\mu}$ ($a^\mu_{\mathbf{k}_\mu}$) for the Bloch state $\psi_{\mathbf{k}_\mu}$. The creation operator for the excitons can be expressed as

$$c^\dagger_{\mathbf{k}_{\mu\nu}}=\int F_{\mathbf{k}_{\mu\nu}}(\mathbf{k})a^{\mu\dagger}_{\mathbf{k}_{\mu\nu}/2+\mathbf{k}}a^{\nu\dagger}_{\mathbf{k}_{\mu\nu}/2-\mathbf{k}}\mathrm{d}\mathbf{k}. \tag{G.41}$$

The commutation relation for exciton creation and annihilation operators is

$$\left[c_{\mathbf{k}'_{\mu\nu}}, c^{\dagger}_{\mathbf{k}_{\mu\nu}}\right] \equiv c_{\mathbf{k}'_{\mu\nu}} c^{\dagger}_{\mathbf{k}_{\mu\nu}} - c^{\dagger}_{\mathbf{k}_{\mu\nu}} c_{\mathbf{k}'_{\mu\nu}} = \delta_{\mathbf{k}_{\mu\nu}\mathbf{k}'_{\mu\nu}} - \Delta_{\mathbf{k}_{\mu\nu}\mathbf{k}'_{\mu\nu}}, \tag{G.42}$$

where the deviation from the Bosonic nature of exciton is

$$\begin{aligned}\Delta_{\mathbf{k}_{\mu\nu}\mathbf{k}'_{\mu\nu}} = &\int F^{*}_{\mathbf{k}'_{\mu\nu}}\left(\mathbf{k}'_{\mu\nu}/2+\mathbf{q}\right) F_{\mathbf{k}_{\mu\nu}}\left(\mathbf{k}_{\mu\nu}/2+\mathbf{q}\right) a^{\mu\dagger}_{\mathbf{k}_{\mu\nu}+\mathbf{q}} a^{\mu}_{\mathbf{k}'_{\mu\nu}+\mathbf{q}} \mathrm{d}\mathbf{q} \\ &+ \int F^{*}_{\mathbf{k}'_{\mu\nu}}\left(-\mathbf{k}_{\mu\nu}/2-\mathbf{q}\right) F_{\mathbf{k}_{\mu\nu}}\left(-\mathbf{k}'_{\mu\nu}/2-\mathbf{q}\right) a^{\nu\dagger}_{\mathbf{k}_{\mu\nu}+\mathbf{q}} a^{\nu}_{\mathbf{k}'_{\mu\nu}+\mathbf{q}} \mathrm{d}\mathbf{q}.\end{aligned} \tag{G.43}$$

The thermal average[5] $\langle \Delta_{\mathbf{k}_{\mu\nu}\mathbf{k}'_{\mu\nu}} \rangle \sim \sum_{\mathbf{k}} \langle a^{\mu\dagger}_{\mathbf{k}} a^{\mu}_{\mathbf{k}} + a^{\mu\dagger}_{\mathbf{k}} a^{\mu}_{\mathbf{k}} \rangle \sim \sum_{\mathbf{k}} \langle c^{\dagger}_{\mathbf{k}} c_{\mathbf{k}} \rangle = n_{\mathrm{x}}$, where n_{x} is the density of the excitons. Therefore, the Bosonic nature of excitons is only approximately valid at low density of excitons.

We have a note here about the sign of the charge carriers in each layers. Consider two piece of identical materials, except that one is composed of anti-matter, i.e., positrons, anti-protons etc. We assume the corresponding particles and anti-particles have same physical properties except their charges are opposite. The electron and positron motions can be described by the same Hamiltonian $H_{\mu} = -\frac{\hbar^2 \nabla^2_{\mathbf{r}_{\mu}}}{2m_{\mu}} + V_{\mu}(\mathbf{r}_{\mu})$ that leads to identical energy band structures and Bloch states. We assume that the charge neutral point (CNP) of the matter and anti-matter materials is at the position such that both are charge neutral if all the states below (above) it are occupied (unoccupied). For the anti-matter material, we assume it is in such an unstable state that only the states with the same quantum numbers as those unoccupied states in the matter material below the CNP are occupied, at any time instant. Now, the electrical conduction in the matter material under external fields can be equally well expressed as the sum of the conduction due to the electrons in the matter material above the CNP and the conduction due to the positrons in the anti-matter material

[5]Here the average $\langle \cdot \rangle \equiv \mathrm{tr}[\rho \cdot]$ where ρ is the density matrix of the system (including all layers and probably the environment).

below the CNP. This decomposition also applies to the interaction with other particles or materials. For example, the Coulomb force exerted on a test charge put in the matter material is equal to the sum of the force between the test charge and the electrons above the CNP and the force between the test charge and the electrons below the CNP plus the back ground charge. The latter is actually equal to the force between the test charge and the positrons states below the CNP. It thus is equally well to describe the conduction and interactions in the matter materials using the electrons above the CNP and positrons below the CNP, as if electrons and positrons can coexist in the matter material (though they cannot in reality).

The inclusion of positrons in our picture is in the same spirit as that of holes in semiconductors. We can now replace the positrons by holes, or more precisely, bare holes since the mass of holes is the same as the mass of electrons. Note that the motions of the electrons and holes are described by the same Hamiltonian. This doesn't mean that they have the same energy and wave function, rather, electrons are restricted above the CNP while holes are below it.

If the materials is doped or gated to be negatively (positively) charged, or it Fermi level is above (below) the CNP, the number of electrons (holes) is much larger than that of holes (electrons), so we can approximately believe that only electrons (holes) are responsible for electromagnetic effects. The accuracy of this description is improved as the temperature drops or if the Fermi level of the matter material lies in the band gap. Thus the sign of the carrier charge in layer μ can be tuned according to the position of the Fermi level. In order for bound state of excitons to exist between layer μ and ν, it is necessary to have one layer hole dominated and the other electron dominated.

We now specifically consider the case that the layer μ and ν have the identical lattice structures, i.e., $V_\mu(\mathbf{r}_\mu) = V_\nu(\mathbf{r}_\nu)$ when $\mathbf{r}_\mu - \mathbf{r}_\nu$ is parallel to the layer normal

and its length is equal to the interlayer distance d. In addition, the bare carrier mass in two layers are the same, i.e., $m_\mu = m_\nu = m$. Thus $V(\mathbf{r}_{\mu\nu}, \bar{\mathbf{r}}_{\mu\nu}) = V_\mu\left(\mathbf{r}_{\mu\nu} + \frac{1}{2}\bar{\mathbf{r}}_{\mu\nu}\right) + V_\nu\left(\mathbf{r}_{\mu\nu} - \frac{1}{2}\bar{\mathbf{r}}_{\mu\nu}\right)$ possesses the same full translation symmetry about $\mathbf{r}_{\mu\nu}$ as each layers, while its Bravais vectors about $\bar{\mathbf{r}}_{\mu\nu}$ are twice longer than and parallel to that about $\mathbf{r}_{\mu\nu}$. Now $\mathbf{r}_{\mu\nu}$ and $\bar{\mathbf{r}}_{\mu\nu}$ can be effectively considered as two-dimensional vectors, by ignoring their components normal to layers. The Coulomb potential can be written as

$$V_{\mu\nu}(\bar{\mathbf{r}}_{\mu\nu}) = \frac{q_\mu q_\nu}{4\pi\epsilon_{\mathrm{r}}\epsilon_0\sqrt{d^2 + \bar{\mathbf{r}}_{\mu\nu}^2}}, \tag{G.44}$$

where ϵ_0 is the vacuum permittivity and ϵ_{r} is the relative permittivity.

We can use plane wave to expand V_μ as

$$V_\mu(\mathbf{r}_\mu) = \sum_{\mathbf{G}_\mu} V_{\mathbf{G}_\mu} e^{i\mathbf{G}_\mu\cdot\mathbf{r}_\mu}. \tag{G.45}$$

Then we have

$$V(\mathbf{r}_{\mu\nu}, \bar{\mathbf{r}}_{\mu\nu}) = \sum_{\mathbf{G}_{\mu\nu}} 2\cos\left(\frac{\mathbf{G}_{\mu\nu}\cdot\bar{\mathbf{r}}_{\mu\nu}}{2}\right) V_{\mathbf{G}_{\mu\nu}} e^{i\mathbf{G}_{\mu\nu}\cdot\mathbf{r}_{\mu\nu}}. \tag{G.46}$$

Thus $V(\mathbf{r}_{\mu\nu}, \bar{\mathbf{r}}_{\mu\nu})$ has the same form as $V_\mu(\mathbf{r}_\mu)$ except that the coefficient of the plane wave is modulated by $2\cos\left(\frac{\mathbf{G}_{\mu\nu}\cdot\bar{\mathbf{r}}_{\mu\nu}}{2}\right)$.

For square lattice with lattice constant a, we can assume $V_{\mathbf{G}_\mu} = v_0$ for $\mathbf{G}_\mu = \pm\mathbf{K}$ and zero otherwise where $\mathbf{K} = \frac{2\pi}{a}(1,1)$, i.e., $V_\mu(\mathbf{r}_\mu) = v_0\cos(\mathbf{K}\cdot\mathbf{r}_\mu)$. The two body potential now becomes

$$V(\mathbf{r}_{\mu\nu}, \bar{\mathbf{r}}_{\mu\nu}) = 2v_0\cos(\mathbf{K}\cdot\mathbf{r}_{\mu\nu})\cos\left(\frac{\mathbf{K}\cdot\bar{\mathbf{r}}_{\mu\nu}}{2}\right) \tag{G.47}$$

which is same as the single layer potential except that the magnitude v_0 is replaced by $2v_0\cos\left(\frac{\mathbf{K}\cdot\bar{\mathbf{r}}_{\mu\nu}}{2}\right)$. If we denote the energy band for layer μ by $E^n_{\mathbf{k}_\mu}(v_0)$ and the Bloch state as

$$\psi_{n\mathbf{k}_\mu} = \sum_{\mathbf{G}_\mu} c^n_{\mathbf{k}_\mu+\mathbf{G}_\mu}(v_0) e^{i(\mathbf{k}_\mu+\mathbf{G}_\mu)\cdot\mathbf{r}_\mu}, \tag{G.48}$$

the Bloch state for the center of motion is

$$\psi_{n\mathbf{k}_{\mu\nu}} = \sum_{\mathbf{G}_{\mu\nu}} c^n_{\mathbf{k}_{\mu\nu}+\mathbf{G}_{\mu\nu}}(2v_0\cos(\mathbf{K}\cdot\bar{\mathbf{r}}_{\mu\nu}/2))e^{i(\mathbf{k}_{\mu\nu}+\mathbf{G}_{\mu\nu})\cdot\mathbf{r}_{\mu\nu}} \tag{G.49}$$

with energy $E_{n\mathbf{k}_\mu}(2v_0\cos(\mathbf{K}\cdot\bar{\mathbf{r}}_{\mu\nu}/2))$. We also have

$$\begin{aligned}\mathbf{A}_{n\mathbf{k}_{\mu\nu}}(\bar{\mathbf{r}}_{\mu\nu}) &= i\hbar\sum_{\mathbf{G}_{\mu\nu}} d^n_{\mathbf{k}_{\mu\nu}+\mathbf{G}_{\mu\nu}}(2v_0\cos(\mathbf{K}\cdot\bar{\mathbf{r}}_{\mu\nu}/2))\nabla_{\bar{\mathbf{r}}_{\mu\nu}}[2v_0\cos(\mathbf{K}\cdot\bar{\mathbf{r}}_{\mu\nu}/2)] \\ &= i\hbar\nabla_{\bar{\mathbf{r}}_{\mu\nu}} D^n_{\mathbf{k}_{\mu\nu}}(2v_0\cos(\mathbf{K}\cdot\bar{\mathbf{r}}_{\mu\nu}/2))\end{aligned} \tag{G.50}$$

where

$$\begin{aligned} d^n_{\mathbf{k}_{\mu\nu}+\mathbf{G}_{\mu\nu}}(x) &= c^{n*}_{\mathbf{k}_{\mu\nu}+\mathbf{G}_{\mu\nu}}(x)\frac{\mathrm{d}}{\mathrm{d}x}c^n_{\mathbf{k}_{\mu\nu}+\mathbf{G}_{\mu\nu}}(x), \\ \frac{\mathrm{d}}{\mathrm{d}x}D^n_{\mathbf{k}_{\mu\nu}}(x) &= \sum_{\mathbf{G}_{\mu\nu}} d^n_{\mathbf{k}_{\mu\nu}+\mathbf{G}_{\mu\nu}}(x).\end{aligned} \tag{G.51}$$

Thus $\mathbf{A}_{n\mathbf{k}_{\mu\nu}}(\bar{\mathbf{r}}_{\mu\nu})$ can be gauged away by the gauge transformation $\psi_{n\mathbf{k}_{\mu\nu}} \to e^{i\theta^n_{\mathbf{k}_{\mu\nu}}}\psi_{n\mathbf{k}_{\mu\nu}}$ where

$$\theta^n_{\mathbf{k}_{\mu\nu}}(\bar{\mathbf{r}}_{\mu\nu}) = iD^n_{\mathbf{k}_{\mu\nu}}(2v_0\cos(\mathbf{K}\cdot\bar{\mathbf{r}}_{\mu\nu}/2)). \tag{G.52}$$

Now the relative motion obeys the following Schroedinger equation

$$\left[-\frac{\hbar^2\nabla^2_{\bar{\mathbf{r}}_{\mu\nu}}}{2m_{\mu\nu}} + V_{\mu\nu}(\bar{\mathbf{r}}_{\mu\nu}) + W_{n\mathbf{k}_{\mu\nu}}(\bar{\mathbf{r}}_{\mu\nu})\right]\phi_{n\mathbf{k}_{\mu\nu}}(\bar{\mathbf{r}}_{\mu\nu}) = e_{\mu\nu}\phi_{n\mathbf{k}_{\mu\nu}}(\bar{\mathbf{r}}_{\mu\nu}), \tag{G.53}$$

where

$$\begin{aligned} W_{n\mathbf{k}_{\mu\nu}}(\bar{\mathbf{r}}_{\mu\nu}) =& \frac{\hbar^2}{2m_{\mu\nu}}\frac{v_0^2 8\pi^2}{a^2}g^n_{\mathbf{k}_{\mu\nu}}(2v_0\cos(\mathbf{K}\cdot\bar{\mathbf{r}}_{\mu\nu}/2))\sin^2\left(\frac{\mathbf{K}\cdot\bar{\mathbf{r}}_{\mu\nu}}{2}\right) \\ &+ E_{n\mathbf{k}_{\mu\nu}}(2v_0\cos(\mathbf{K}\cdot\bar{\mathbf{r}}_{\mu\nu}/2)),\end{aligned} \tag{G.54}$$

with

$$g^n_{\mathbf{k}_{\mu\nu}}(x) = \sum_{\mathbf{G}_{\mu\nu}}\left[\frac{\mathrm{d}}{\mathrm{d}x}c^{n*}_{\mathbf{k}_{\mu\nu}+\mathbf{G}_{\mu\nu}}(x)\frac{\mathrm{d}}{\mathrm{d}x}c^n_{\mathbf{k}_{\mu\nu}+\mathbf{G}_{\mu\nu}}(x)\right] + \left[\sum_{\mathbf{G}_{\mu\nu}} d^n_{\mathbf{k}_{\mu\nu}+\mathbf{G}_{\mu\nu}}(x)\right]^2. \tag{G.55}$$

If layer μ and ν have opposite sign of carrier charge, the relative motion is in the off-plane hydrogenic potential $V_{\mu\nu}$ augmented by a periodic potential $W_{n\mathbf{k}_{\mu\nu}}(\bar{\mathbf{r}}_{\mu\nu})$. It is proved [202] that such system permits at least one bound state with energy below the minimum of the energy band produced by the periodic potential $W_{n\mathbf{k}_{\mu\nu}}(\bar{\mathbf{r}}_{\mu\nu})$.

When uniform external fields $\mathbf{E}_\mu$ are applied to each layer, the two body Hamiltonian is modified to

$$H_{\mu\nu} - q_\mu \mathbf{E}_\mu \cdot \mathbf{r}_\mu - q_\nu \mathbf{E}_\nu \cdot \mathbf{r}_\nu = H_{\mu\nu} - (q_\mu \mathbf{E}_\mu + q_\nu \mathbf{E}_\nu) \cdot \mathbf{r}_{\mu\nu} - \frac{1}{2}(q_\mu \mathbf{E}_\mu - q_\nu \mathbf{E}_\nu) \cdot \bar{\mathbf{r}}_{\mu\nu}. \quad \text{(G.56)}$$

Thus, the center of mass motion for excitons is influenced by a force of $q_\mu \mathbf{E}_\mu + q_\nu \mathbf{E}_\nu$, while the internal structure of the exciton is modified by the presence of the potential $-\frac{1}{2}(q_\mu \mathbf{E}_\mu - q_\nu \mathbf{E}_\nu) \cdot \bar{\mathbf{r}}_{\mu\nu}$. Since the external fields can be arbitrarily small, we can ignore the change of the internal structures of excitons due to external fields. Therefore, the semi-classical dynamics of excitons is

$$\begin{aligned} \partial_t \mathbf{k}_{\mu\nu} &= \frac{1}{\hbar}(\mathbf{F}_\mu + \mathbf{F}_\nu) = \frac{1}{\hbar}(q_\mu \mathbf{E}_\mu + q_\nu \mathbf{E}_\nu), \\ \partial_t \mathbf{x} &= \mathbf{v}_{\mu\nu}(\mathbf{k}_{\mu\nu}) = \frac{1}{\hbar}\nabla_{\mathbf{k}} e_{\mu\nu}(\mathbf{k}_{\mu\nu}). \end{aligned} \quad \text{(G.57)}$$

We ignore the Berry phase effect of excitons here.

G.4 Linearizing Boltzmann equation for excitons

The collision integrals with excitons involved are defined in the following:

$$
\begin{aligned}
D_{\mu\nu}[f_\mu, f_\nu, f_{\mu\nu}](\mathbf{k}_\mu) &= \int_{B_\nu B_{\mu\nu}} [S_{\mathrm{xca}}(\mathbf{x}, \mathbf{k}_{\mu\nu}, \mathbf{k}_\mu, \mathbf{k}_\nu) f_{\mu\nu} \bar{f}_\mu \bar{f}_\nu \\
&\qquad - S_{\mathrm{xca}}(\mathbf{x}, \mathbf{k}_\mu, \mathbf{k}_\nu, \mathbf{k}_{\mu\nu}) f_\mu f_\nu \bar{f}_{\mu\nu}] \mathrm{d}\mathbf{k}_\nu \mathrm{d}\mathbf{k}_{\mu\nu}, \\
D^{\mathrm{x}}_{\mu\nu}[f_\mu, f_\nu, f_{\mu\nu}](\mathbf{k}_{\mu\nu}) &= \int_{B_\mu B_\nu} [S_{\mathrm{xca}}(\mathbf{x}, \mathbf{k}_\mu, \mathbf{k}_\nu, \mathbf{k}_{\mu\nu}) f_\mu f_\nu \bar{f}_{\mu\nu} \\
&\qquad - S_{\mathrm{xca}}(\mathbf{x}, \mathbf{k}_{\mu\nu}, \mathbf{k}_\mu, \mathbf{k}_\nu) f_{\mu\nu} \bar{f}_\mu \bar{f}_\nu] \mathrm{d}\mathbf{k}_\mu \mathrm{d}\mathbf{k}_\nu, \\
C^{\mathrm{x}}_{\mu\nu}[f_{\mu\nu}](\mathbf{k}_{\mu\nu}) &= \int_{B_{\mu\nu}} [S_{\mathrm{x}}(\mathbf{x}, \mathbf{k}'_{\mu\nu}, \mathbf{k}_{\mu\nu}) f'_{\mu\nu} \bar{f}_{\mu\nu} - S_{\mathrm{x}}(\mathbf{x}, \mathbf{k}_{\mu\nu}, \mathbf{k}'_{\mu\nu}) f_{\mu\nu} \bar{f}'_{\mu\nu}] \mathrm{d}\mathbf{k}'_{\mu\nu}, \\
C^{\mathrm{x}}_{\mu\nu,\alpha\beta}[f_{\mu\nu}, f_{\alpha\beta}](\mathbf{k}_{\mu\nu}) &= \int_{B_{\mu\nu} B^2_{\alpha\beta}} [S_{\mathrm{xx}}(\mathbf{x}, \mathbf{k}'_{\mu\nu}, \mathbf{k}_{\mu\nu}, \mathbf{k}'_{\alpha\beta}, \mathbf{k}_{\alpha\beta}) f'_{\mu\nu} \bar{f}_{\mu\nu} f'_{\alpha\beta} \bar{f}_{\alpha\beta} \\
&\qquad - S_{\mathrm{xx}}(\mathbf{x}, \mathbf{k}_{\mu\nu}, \mathbf{k}'_{\mu\nu}, \mathbf{k}_{\alpha\beta}, \mathbf{k}'_{\alpha\beta}) f_{\mu\nu} \bar{f}'_{\mu\nu} f_{\alpha\beta} \bar{f}'_{\alpha\beta}] \mathrm{d}\mathbf{k}'_{\mu\nu} \mathrm{d}\mathbf{k}_{\alpha\beta} \mathrm{d}\mathbf{k}'_{\alpha\beta}, \\
D_{\alpha\beta,\mu}[f_{\alpha\beta}, f_\mu](\mathbf{k}_\mu) &= \int_{B^2_{\alpha\beta} B_\mu} [S_{\mathrm{xc}}(\mathbf{x}, \mathbf{k}'_{\alpha\beta}, \mathbf{k}_{\alpha\beta}, \mathbf{k}'_\mu, \mathbf{k}_\mu) f'_{\alpha\beta} \bar{f}_{\alpha\beta} f'_\mu \bar{f}_\mu \\
&\qquad - S_{\mathrm{xc}}(\mathbf{x}, \mathbf{k}_{\alpha\beta}, \mathbf{k}'_{\alpha\beta}, \mathbf{k}_\mu, \mathbf{k}'_\mu) f_{\alpha\beta} \bar{f}'_{\alpha\beta} f_\mu \bar{f}'_\mu] \mathrm{d}\mathbf{k}_{\alpha\beta} \mathrm{d}\mathbf{k}'_{\alpha\beta} \mathrm{d}\mathbf{k}'_\mu, \\
D^{\mathrm{x}}_{\mu\nu,\alpha}[f_{\mu\nu}, f_\alpha](\mathbf{k}_{\mu\nu}) &= \int_{B_{\mu\nu} B^2_\alpha} [S_{\mathrm{xc}}(\mathbf{x}, \mathbf{k}'_{\mu\nu}, \mathbf{k}_{\mu\nu}, \mathbf{k}'_\alpha, \mathbf{k}_\alpha) f'_{\mu\nu} \bar{f}_{\mu\nu} f'_\alpha \bar{f}_\alpha \\
&\qquad - S_{\mathrm{xc}}(\mathbf{x}, \mathbf{k}_{\mu\nu}, \mathbf{k}'_{\mu\nu}, \mathbf{k}_\alpha, \mathbf{k}'_\alpha) f_{\mu\nu} \bar{f}'_{\mu\nu} f_\alpha \bar{f}'_\alpha] \mathrm{d}\mathbf{k}'_{\mu\nu} \mathrm{d}\mathbf{k}_\alpha \mathrm{d}\mathbf{k}'_\alpha,
\end{aligned}
\tag{G.58}
$$

where $\bar{f}_{\mu\nu} \equiv 1 + f_{\mu\nu}$ since excitons are considered to be bosons. It can be easily checked that

$$\int_{B_\mu} \left(C_\mu[f_\mu] + \sum_\nu C_{\mu\nu}[f_\mu, f_\nu] + \sum_\nu D_{\mu\nu}[f_\mu, f_\nu, f_{\mu\nu}] + \sum_{\alpha\beta} D_{\alpha\beta,\mu}[f_{\alpha\beta}, f_\mu] \right) \mathrm{d}\mathbf{k}_\mu$$
$$+ \sum_\nu \int_{B_{\mu\nu}} \left(C^{\mathrm{x}}_{\mu\nu}[f_{\mu\nu}] + D^{\mathrm{x}}_{\mu\nu}[f_\mu, f_\nu, f_{\mu\nu}] + \sum_{\alpha\beta} C^{\mathrm{x}}_{\mu\nu,\alpha\beta}[f_{\mu\nu}, f_{\alpha\beta}] \right.$$
$$\left. + \sum_\alpha D^{\mathrm{x}}_{\mu\nu,\alpha}[f_{\mu\nu}, f_\alpha] \right) \mathrm{d}\mathbf{k}_{\mu\nu} = 0 \tag{G.59}$$

for any given μ, thus the carrier number of each layer is conserved. These collision integrals describes the exciton creations and annihilations ($D_{\mu\nu}$ and $D^{\mathrm{x}}_{\mu\nu}$), single exciton scatterings ($C^{\mathrm{x}}_{\mu\nu}$, due to phonons, defects, impurity etc.), exciton-exciton scatterings ($C^{\mathrm{x}}_{\mu\nu,\alpha\beta}$) and exciton-carrier scatterings ($D_{\alpha\beta,\mu}$ and $D^{\mathrm{x}}_{\mu\nu,\alpha}$), and their corresponding transition rates are S_{xca}, S_{x}, S_{xx} and S_{xc}, respectively. The other conservation laws (energy, momentum etc.) are encoded in the transition rates.

The equilibrium distribution of excitons are

$$n_{\mu\nu}(\mathbf{k}_{\mu\nu}) = \left[\exp\left(\frac{e_{\mu\nu}(\mathbf{k}_{\mu\nu}) - \zeta_{\mu\nu}}{k_{\mathrm{B}}T} \right) - 1 \right]^{-1} \tag{G.60}$$

where $\zeta_{\mu\nu}$ is the chemical potential. The above collision integrals are linearized by writing

$$f_{\mu\nu} = n_{\mu\nu}(1 - \bar{n}_{\mu\nu}\phi_{\mu\nu}). \tag{G.61}$$

We consider the spatially homogeneous system, and thus the position dependence is dropped. The linearized collision integrals are

$$
\begin{aligned}
D_{\mu\nu}[\phi_\mu,\phi_\nu,\phi_{\mu\nu}](\mathbf{k}_\mu) &= \int_{B_\nu B_{\mu\nu}} Q_{\mathrm{xca}}([\mu\nu]\mu\nu)[\phi_\mu+\phi_\nu-\phi_{\mu\nu}]\mathrm{d}\nu\mathrm{d}[\mu\nu],\\
D^{\mathrm{x}}_{\mu\nu}[\phi_\mu,\phi_\nu,\phi_{\mu\nu}](\mathbf{k}_{\mu\nu}) &= -\int_{B_\mu B_\nu} Q_{\mathrm{xca}}([\mu\nu]\mu\nu)[\phi_\mu+\phi_\nu-\phi_{\mu\nu}]\mathrm{d}\mu\mathrm{d}\nu,\\
C^{\mathrm{x}}_{\mu\nu}[\phi_{\mu\nu}](\mathbf{k}_{\mu\nu}) &= \int_{B_{\mu\nu}} Q_{\mathrm{x}}([\mu\nu]'[\mu\nu])[\phi_{\mu\nu}-\phi'_{\mu\nu}]\mathrm{d}[\mu\nu]',\\
C^{\mathrm{x}}_{\mu\nu,\alpha\beta}[\phi_{\mu\nu},\phi_{\alpha\beta}](\mathbf{k}_{\mu\nu}) &= \int_{B_{\mu\nu}B^2_{\alpha\beta}} Q_{\mathrm{xx}}([\mu\nu]'[\mu\nu][\alpha\beta]'[\alpha\beta])\times\\
&\qquad [\phi_{\mu\nu}-\phi'_{\mu\nu}+\phi_{\alpha\beta}-\phi'_{\alpha\beta}]\mathrm{d}[\mu\nu]'\mathrm{d}[\alpha\beta]\mathrm{d}[\alpha\beta]',\\
D_{\alpha\beta,\mu}[\phi_{\alpha\beta},\phi_\mu](\mathbf{k}_\mu) &= \int_{B^2_{\alpha\beta}B_\mu} Q_{\mathrm{xc}}([\alpha\beta]'[\alpha\beta]\mu'\mu)\times\\
&\qquad [\phi_\mu-\phi'_\mu+\phi_{\alpha\beta}-\phi'_{\alpha\beta}]\mathrm{d}\mu'\mathrm{d}[\alpha\beta]\mathrm{d}[\alpha\beta]',\\
D^{\mathrm{x}}_{\mu\nu,\alpha}[\phi_{\mu\nu},\phi_\alpha](\mathbf{k}_{\mu\nu}) &= \int_{B_{\mu\nu}B^2_\alpha} Q_{\mathrm{xc}}([\mu\nu]'[\mu\nu]\alpha'\alpha)\times\\
&\qquad [\phi_{\mu\nu}-\phi'_{\mu\nu}+\phi_\alpha-\phi'_\alpha]\mathrm{d}[\mu\nu]'\mathrm{d}\alpha\mathrm{d}\alpha',
\end{aligned}
\tag{G.62}
$$

where

$$
\begin{aligned}
Q_{\mathrm{xca}}([\mu\nu]\mu\nu) &\equiv S_{\mathrm{xca}}(\mathbf{x},\mathbf{k}_{\mu\nu},\mathbf{k}_\mu,\mathbf{k}_\nu)n_{\mu\nu}\bar{n}_\mu\bar{n}_\nu,\\
Q_{\mathrm{x}}([\mu\nu]'[\mu\nu]) &\equiv S_{\mathrm{x}}(\mathbf{k}'_{\mu\nu},\mathbf{k}_{\mu\nu})n'_{\mu\nu}\bar{n}_{\mu\nu},\\
Q_{\mathrm{xx}}([\mu\nu]'[\mu\nu][\alpha\beta]'[\alpha\beta]) &\equiv S_{\mathrm{xx}}(\mathbf{k}'_{\mu\nu},\mathbf{k}_{\mu\nu},\mathbf{k}'_{\alpha\beta},\mathbf{k}_{\alpha\beta})n'_{\mu\nu}\bar{n}_{\mu\nu}n'_{\alpha\beta}\bar{n}_{\alpha\beta},\\
Q_{\mathrm{xc}}([\alpha\beta]'[\alpha\beta]\mu'\mu) &\equiv S_{\mathrm{xc}}(\mathbf{k}'_{\alpha\beta},\mathbf{k}_{\alpha\beta},\mathbf{k}'_\mu,\mathbf{k}_\mu)n'_{\alpha\beta}\bar{n}_{\alpha\beta}n'_\mu\bar{n}_\mu,\\
\mathrm{d}[\mu\nu] &\equiv \mathrm{d}\mathbf{k}_{\mu\nu}
\end{aligned}
\tag{G.63}
$$

and $[\mu\nu]$ is used to denote $\mathbf{k}_{\mu\nu}$. The principle of detailed balance requires

$$\begin{aligned} Q_{\mathrm{xca}}([\mu\nu]\mu\nu) &= Q_{\mathrm{xca}}(\mu\nu[\mu\nu]), \\ Q_{\mathrm{x}}([\mu\nu]'[\mu\nu]) &= Q_{\mathrm{x}}([\mu\nu][\mu\nu]'), \\ Q_{\mathrm{xx}}([\mu\nu]'[\mu\nu][\alpha\beta]'[\alpha\beta]) &= Q_{\mathrm{xx}}([\mu\nu][\mu\nu]'[\alpha\beta][\alpha\beta]'), \\ Q_{\mathrm{xc}}([\alpha\beta]'[\alpha\beta]\mu'\mu) &= Q_{\mathrm{xc}}([\alpha\beta][\alpha\beta]'\mu\mu'). \end{aligned} \tag{G.64}$$

Additional symmetries are

$$\begin{aligned} Q_{\mathrm{xca}}([\mu\nu]\mu\nu) &= Q_{\mathrm{xca}}([\mu\nu]\nu\mu), \\ Q_{\mathrm{xx}}([\mu\nu]'[\mu\nu][\alpha\beta]'[\alpha\beta]) &= Q_{\mathrm{xx}}([\alpha\beta]'[\alpha\beta][\mu\nu]'[\mu\nu]), \\ Q_{\mathrm{xc}}([\alpha\beta]'[\alpha\beta]\mu'\mu) &= Q_{\mathrm{xc}}(\mu'\mu[\alpha\beta]'[\alpha\beta]), \end{aligned} \tag{G.65}$$

and the layer indices inside the square bracket can be freely exchanged since $f_{\mu\nu} = f_{\nu\mu}$. For every pair indices for excitons such as $\mu\nu$, it implicitly requires $\mu \neq \nu$. They can be identical for carriers in each layer.

G.4.1 Linear operators

They are separated into four groups.

- First type:

$$\begin{aligned} P_{\mu\mu,\nu\nu}\phi_\nu =& \delta_{\mu\nu}\left[\int_{B_\mu} W(\mu'\mu)(\phi_\mu - \phi'_\mu)\mathrm{d}\mu' + G(\mu)\phi_\mu\right] + \int_{B_\nu} g(\mu\nu)\phi_\nu \mathrm{d}\nu \\ &+ \int_{B_\mu B_\nu^2} Q_{\mathrm{cc}}(\mu'\mu\nu'\nu)(\phi_\nu - \phi'_\nu)\mathrm{d}\mu'\mathrm{d}\nu\mathrm{d}\nu', \end{aligned} \tag{G.66}$$

where

$$\begin{aligned} W(\mu'\mu) &\equiv Q_{\rm c}(\mu'\mu) + \sum_{\alpha}\int_{B^2_\alpha} Q_{\rm cc}(\mu'\mu\alpha'\alpha){\rm d}\alpha{\rm d}\alpha' \\ &\quad + \sum_{\alpha\neq\beta}\int_{B^2_{\alpha\beta}} Q_{\rm xc}([\alpha\beta]'[\alpha\beta]\mu'\mu){\rm d}[\alpha\beta]{\rm d}[\alpha\beta]', \\ G(\mu) &\equiv \sum_{\alpha\neq\mu}\int_{B_{\mu\alpha}B_\alpha} Q_{\rm xca}([\mu\alpha]\mu\alpha){\rm d}[\mu\alpha]{\rm d}\alpha, \\ g(\mu\nu) &= \int_{B_{\mu\nu}} Q_{\rm xca}([\mu\nu]\mu\nu){\rm d}[\mu\nu]. \end{aligned} \tag{G.67}$$

Apparently we have $W(\mu'\mu) = W(\mu\mu')$ and $g(\mu\nu) = g(\nu\mu)$.

- Second type ($\alpha \neq \beta$):

$$\begin{aligned} P_{\mu\mu,\alpha\beta}\phi_{\alpha\beta} &= \int_{B^2_{\alpha\beta}B_\mu} Q_{\rm xc}([\alpha\beta]'[\alpha\beta]\mu'\mu)(\phi_{\alpha\beta} - \phi'_{\alpha\beta}){\rm d}[\alpha\beta]{\rm d}[\alpha\beta]'{\rm d}\mu' \\ &\quad - \delta_{\mu\alpha}\int_{B_{\alpha\beta}B_\alpha} Q_{\rm xca}([\alpha\beta]\alpha\beta)\phi_{\alpha\beta}{\rm d}[\alpha\beta]{\rm d}\beta. \end{aligned} \tag{G.68}$$

- Third type ($\mu \neq \nu$):

$$\begin{aligned} P_{\mu\nu,\alpha\alpha}\phi_\alpha &= \int_{B_{\mu\nu}B^2_\alpha} Q_{\rm xc}([\mu\nu]'[\mu\nu]\alpha'\alpha)(\phi_\alpha - \phi'_\alpha){\rm d}[\mu\nu]'{\rm d}\alpha{\rm d}\alpha' \\ &\quad - \int_{B_\alpha} h([\mu\nu]\alpha)\phi_\alpha{\rm d}\alpha, \end{aligned} \tag{G.69}$$

where

$$h([\mu\nu]\alpha) \equiv \delta_{\alpha\mu}\int_{B_\nu} Q_{\rm xca}([\mu\nu]\alpha\nu){\rm d}\nu + \delta_{\alpha\nu}\int_{B_\mu} Q_{\rm xca}([\mu\nu]\mu\alpha){\rm d}\mu. \tag{G.70}$$

- Fourth type ($\mu \neq \nu$, $\alpha \neq \beta$):

$$\begin{aligned} P_{\mu\nu,\alpha\beta}\phi_{\alpha\beta} =& \delta_{\mu\nu,\alpha\beta}\left[\int_{B_{\mu\nu}} M([\mu\nu]'[\mu\nu])(\phi_{\mu\nu} - \phi'_{\mu\nu}){\rm d}[\mu\nu]' + H([\mu\nu])\phi_{\mu\nu}\right] \\ &+ \int_{B_{\mu\nu}B^2_{\alpha\beta}} Q_{\rm xx}([\mu\nu]'[\mu\nu][\alpha\beta]'[\alpha\beta])(\phi_{\alpha\beta} - \phi'_{\alpha\beta}){\rm d}[\mu\nu]'{\rm d}[\alpha\beta]{\rm d}[\alpha\beta]', \end{aligned} \tag{G.71}$$

where

$$M([\mu\nu]'[\mu\nu]) \equiv Q_{\mathrm{x}}([\mu\nu]'[\mu\nu]) + \sum_{\rho} \int_{B_{\rho}^{2}} Q_{\mathrm{xc}}([\mu\nu]'[\mu\nu]\rho'\rho)\mathrm{d}\rho\mathrm{d}\rho' + \sum_{\rho\gamma} \int_{B_{\rho\gamma}^{2}} Q_{\mathrm{xx}}([\mu\nu]'[\mu\nu][\rho\gamma]'[\rho\gamma])\mathrm{d}[\rho\gamma]\mathrm{d}[\rho\gamma]', \tag{G.72}$$

$$H([\mu\nu]) \equiv \int_{B_{\mu}B_{\nu}} Q_{\mathrm{xca}}([\mu\nu]\mu\nu)\mathrm{d}\mu\mathrm{d}\nu.$$

For any given ψ_{μ} and ϕ_{ν}, we have

$$\langle \psi_{\mu}, P_{\mu\mu,\nu\nu}\phi_{\nu} \rangle = \langle \phi_{\nu}, P_{\nu\nu,\mu\mu}\psi_{\mu} \rangle , \tag{G.73}$$

for the first type of operators and similar for any given $\psi_{\mu\nu}$ and $\phi_{\alpha\beta}$, we have

$$\langle \psi_{\mu\nu}, P_{\mu\nu,\alpha\beta}\phi_{\alpha\beta} \rangle = \langle \phi_{\alpha\beta}, P_{\alpha\beta,\mu\nu}\psi_{\mu\nu} \rangle , \tag{G.74}$$

for the fourth type of operators.

VITA

VITA

CONTACTS

E-mails: hujiuning@gmail.com, hu49@purdue.edu

Web: http://sites.google.com/site/hujiuning

EDUCATION

PhD (2015), School of Electrical and Computer Engineering, Purdue University

M.S. (2008), Institute of Microelectronics, Tsinghua University

B.S. (2006), Department of Electronic Engineering, Tsinghua University

SELECTED PUBLICATIONS

- Y. Wang, A. Vallabhaneni and **J. Hu** *et al.*, *Nano Lett.*, 14, 592 (2014)
- **J. Hu** *et al.*, *ECS Transactions*, 53, 41 (2013)
- **J. Hu** and Y. Chen, *Phys. Rev. E*, 87, 062104 (2013)
- A. Vallabhaneni, B. Qiu and **J. Hu** *et al.*, *J. Appl. Phys.*, 113, 064311 (2013)
- K. Gunawardana, K. Mullen and **J. Hu** *et al.*, *Phys. Rev. B*, 85, 245417 (2012)
- **J. Hu** *et al.*, *Mater. Res. Soc. Symp. Proc.*, 1456-JJ05-05 (2012)
- **J. Hu** *et al.*, *International Journal of Thermophysics*, 33, 986 (2012)
- **J. Hu** *et al.*, *Appl. Phys. Lett.*, 99, 113101 (2011)
- **J. Hu** *et al.*, *Appl. Phys. Lett.*, 97, 133107 (2010)
- **J. Hu** *et al.*, *Nano Lett.*, 9, 2730 (2009)
- **J. Hu** *et al.*, *AIP Conf. Proc.*, 1173, 135 (2009)
- **J. Hu** *et al.*, *Mater. Res. Soc. Symp. Proc.*, 0998-J02-03 (2007)
- M. Ren, L. Zhang and **J. Hu** *et al.*, *Acta Physica Sinica*, 56, 2863-2867 (2007)